DOCUMENT RESUME

ED 223 552	SP 021 293
AUTHOR	Dickson, George E.; Wiersma, William
TITLE	Measurement of Teacher Competence. Research and Evaluation in Teacher Education.
INSTITUTION	Toledo Univ., Ohio. Center for Educational Research and Services.
PUB DATE	Oct 82
NOTE	117p.; For related document, see ED 194 540.
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	*Classroom Observation Techniques; *Competency Based Teacher Education; *Evaluation Methods; Higher Education; Preservice Teacher Education; *Program Evaluation; *Research Methodology; Student Teachers; Student Teaching; Teacher Behavior; Teacher Education Programs; Teacher Effectiveness; *Teacher Evaluation
IDENTIFIERS	*University of Toledo OH

ABSTRACT

This monograph contains a description of continued research efforts in measuring and evaluating preservice teacher competencies and effectiveness in the competency-based teacher education (CBTE) program at the University of Toledo (Ohio). Chapter 1 introduces the assessment rationale, and chapter 2 provides an overview of research literature on teacher effectiveness. Chapter 3 focuses on the identification of 49 teaching competencies and the methodology of classroom or teacher observation used at the University of Toledo. This chapter also includes a discussion of high and low inference observation and a description of the two measurement instruments: Classroom Observations Keyed for Effectiveness Research (COKER) and Teacher Performance Assessment Instruments (TPAI). Chapters 4 and 5 provide the results of two different types of data analysis--profile analysis and multivariate analysis of TPAI and COKER data. The final chapter presents conclusions, makes recommendations for further action, and outlines the uses of the data. (JD)

****
*   Reproductions supplied by EDRS are the best that can be made   *
*   from the original document.   *
****

ED223552

RESEARCH AND EVALUATION IN TEACHER EDUCATION

MEASUREMENT OF TEACHER COMPETENCE

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
George E. Dickson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

by

George E. Dickson
William Wiersma

The Center for Educational Research and Services
College of Education and Allied Professions
The University of Toledo
Toledo, Ohio

October 1982

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☑ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

SP021293

PREFACE

This monograph reports a continued effort at The University of Toledo, Toledo, Ohio, U.S.A., to conduct research and evaluation of an operational, competency-based teacher education (CBTE) program. Teacher education continues to receive, for the most part, negative comment from the general public. Assumptions are constantly made that teachers are unprepared in subject matter and general teaching skills. Any evidence to the contrary is brushed aside and the negative tends to be rather constantly portrayed. As we said in our previous report*, there has been considerable change in the 1960's and '70's in teacher education, much of it directed at curriculum. At the same time, teacher education as a discipline experiences relatively little in the way of empirical investigation. Research to measure and to evaluate teaching performance on the basis of training programs is more of a goal than a realized process, with widely reported results. Nevertheless, efforts are being made, worldwide, to extend such research and obtain valid data on the measurement of teacher competencies.

The University of Toledo embarked on a measurement and evaluation effort for its CBTE program in 1975. As the process evolved, a procedure and instruments have been developed and used in an attempt to measure teacher competence and teacher performance with the view of also considering pupil learning experiences and outcomes resulting from such performance. Our initial report on this effort appeared in a monograph released in May 1980. We now have further data analyzed and certain refinements of our previous research efforts in place. This monograph contains a description of our continued research efforts in the area of teacher competence and effectiveness from the CBTE perspective at The University of Toledo.

The monograph consists of six chapters. Chapter I introduces the context of the monograph. Chapter II provides a limited review of relevant literature on the subject. Chapter III indicates procedures for identifying teacher competencies and the classroom observation methods and instruments to do so. Chapters IV and V provide two different types of data analysis — profile analysis and multivariate

---

*George E. Dickson and William Wiersma, *Research and Evaluation in Teacher Education: A Concern for Competent, Effective Teachers*, Toledo: The Center for Educational Research and Services, College of Education, The University of Toledo, May 1980. (Also in ERIC) analysis. The final chapter, VI, establishes conclusions, makes recommendations for further action, and considers where the topic being explored is taking the College of Education and Allied Professions at The University of Toledo as well as other teacher education institutions and the profession at large.

ACKNOWLEDGEMENTS

A research/evaluation effort to measure teacher competence requires the cooperation and assistance of many individuals. George E. Dickson has served as Project Director and an able research team has been organized to conduct the extensive activities required by the project. The team in operation during the time period covered in this report 1980-82 consisted of the following individuals.

Dr. William Wiersma served as the project Co-Director and provided special assistance in data preparation and analysis, as well as the development of the project research design and procedures. Dr. Stephen Jurs provided assistance in factor analysis and Dr. Judith Johnson-Wenig developed the computer programs and analysis procedures utilized in the study.

The collection of data on the COKER and the TPAI is coordinated through the help of two faculty members. Dr. John F. Schaff coordinated data collection for the COKER and Dr. George B. Shirk coordinated data collection for the TPAI.

The data collection team on the COKER consisted of twelve faculty members who were specially trained for this activity. These persons were Dr. David Balzer, Dr. Langston Bannister, Dr. Keith Bernhard, Dr. Jerry DeBruin, Dr. James Gress, Dr. Dwayne DeMedio, Dr. Lester Elsie, Dr. Alice Kildea, Ms. Judith Scheer, Dr. Cynthia Warger, Dr. Robert Wilhoyte, and Dr. Darryl Yorke.

The data collectors for the TPAI consisted of faculty members who regularly supervised student teachers as well as graduate assistants who had student teaching supervisory roles assigned to them. This group of people was quite large and involved the following individuals: Dr. John Ahern, Dr. Charles Armstrong, Dr. Alan Ashby, Ms. Barbara Baldwin, Dr. David Balzer, Dr. Langston Bannister, Dr. Reemt Baumann, Ms. Cynthia Brady, Mr. Merrill Brown, Mr. Gary Cashin, Ms. Elizabeth Cole, Dr. Gary Cooke, Dr. John Cryan, Dr. Dewitt Davison, Dr. Jerome DeBruin, Dr. Dwayne DeMedio, Dr. Clyde Dilley, Ms. Diane Dixon, Dr. Diane DuBois, Ms. Linda Dugger, Dr. Lester Elsie, Dr. Thomas Gibney, Dr. George Gilmore, Ms. Helen Glanville, Dr. William Gray, Dr. James Gress, Ms. Sandra Harding, Dr. Lois Hodgson, Dr. Sally Holt, Ms. Bonita Hudak, Dr. Lynne Hudson, Mr. Theodore Johnson, Dr. Alice Kildea, Mrs. Mary Louise Krill, Mrs. Phyllis Lamb, Dr. Thomas Lopez, Ms. Ruth Lindecker, Dr. Ralph

Martin, Dr. Dean Meinke, Dr. Hughes Moir, Dr. Dennis Myers, Ms. Anthula Natsoulas, Ms. Kathleen Nejman, Dr. Guessler Normand, Mrs. Janet Northrup, Mr. Douglas Nichol, Mr. Lawrence Orrach, Dr. Judythe Patberg, Dr. Judith Rink, Mrs. Jerilyn Riordan, Dr. Herbert Sandberg, Dr. John Schaff, Mrs. Judith Scheer, Dr. George Shirk, Dr. Donald Small, Dr. Sam Snyder, Dr. John Stammer, Dr. Donald Stolberg, Dr. Jerry Underfer, Dr. Cynthia Warger, Dr. Robert Wilhoyte, and Dr. Darryl Yorke.

A special recognition is due for four faculty members who provided information and copy for Chapter 3 as well as help in dissemination efforts conducted at two national meetings (AACTE and ATE) in the Spring of 1982. These individuals were Dr. David Balzer, Dr. Jerry DeBruin, Dr. Dwayne DeMedio, and Dr. Lester Elsie. Their assistance in helping prepare Chapter 3 is very much appreciated.

Acknowledgement of persons helpful in preparing this report must also include the expert secretarial service and graphics presentation help of Mrs. Terry Bush and Mrs. Mary Garrett.

The authors of this manuscript are much indebted to the above persons who have all been instrumental in varying degrees to making a difficult project possible.

George E. Dickson
William Wiersma

CONTENTS

Preface

Acknowledgements

I. AN INTRODUCTION ... 1

II. A LIMITED REVIEW OF RELEVANT LITERATURE ... 8

The General Concern for Competent, Effective Teachers ... 8
A General Review of Research on Teacher Competency, Performance and Effectiveness ... 9

III. THE IDENTIFICATION OF TEACHING COMPETENCIES AND THE METHODOLOGY OF CLASSROOM OBSERVATION ... 23

The Identification of Teaching Competencies ... 23
  Description of The University of Toledo CBTE Program ... 23
  Competencies Based on Content Analysis ... 24
The Methodology of Classroom Observation ... 26
  Low-Inference Observation ... 26
    The COKER as a Low-Inference Instrument ... 28
  High-Inference Observation ... 31
    The TPAI as a High-Inference Instrument ... 32
The Measurement of Competencies ... 36
  Mapping the Behaviors Into the Competencies Measured by the Instruments ... 36
  The Scoring of COKER and TPAI Data ... 37
  The Nature of Competency Scores ... 39
Validity and Reliability ... 41
  Reliability of the TPAI ... 43
  Reliability of the COKER ... 44
  Summary ... 44

IV. PROFILE ANALYSIS AND RESULTS ... 46

Profile Analysis ... 46
  TPAI Results ... 46
  COKER Results ... 53
Summary ... 59

V. MULTIVARIATE ANALYSIS OF OBSERVATION DATA ... 62

Factor Analysis ... 62
  TPAI Results ... 63
  COKER Results ... 67
  Comparison of TPAI and COKER Factor Patterns ... 67
Multivariate Analyses of Variance ... 71
  TPAI Results ... 72
  COKER Results ... 74
Comparison of Factor Analysis and MANOVA ... 75
Summary ... 76

VI. OTHER RESULTS, CONSIDERATIONS AND CONCLUSIONS ... 78
Competencies Measured by Both the COKER and TPAI ... 78
The Use of Observation Data In Program Evaluation ... 82
Individual Diagnosis and Assessment ... 84
Other Uses of Observation Data ... 88
  Assessment of Classroom Interaction ... 92
  Inservice and Staff Development Programs ... 92
  Certification of Teachers ... 93
Conclusions and Recommendations ... 93

Appendix A. University of Toledo Competencies Assessed Using the TPAI and COKER Instruments and Competency Symbol Indicating Such Assessment ... 101
Appendix B. The Generic Competencies of Effective Teaching as Related to the Ohio Standards for Colleges and Universities Preparing Teachers ... 105
Appendix C. Classroom Observations Keyed for Effectiveness Research ... 108

Chapter I

AN INTRODUCTION

The development and use of high and low inference observation[1] of teacher behavior in the classroom as well as other procedures to measure teacher competence, performance and effectiveness continue to be a major concern of teacher educators, other education professionals and the public. There continues to be an increasing demand for the use of teacher assessment procedures as witnessed by the growing number of states, state education agencies and others in the development and use of assessment procedures.[2]

Resistance to research, evaluation, or assessment of teacher competence continues to come from various segments of the education profession and other societal groups usually on the argument that measurement procedures are improper and/or the results inconclusive. However, regardless of such criticism, the national drive for educational accountability through demonstrated teacher competence continues and requires constant attention as to whether teacher training programs purporting to develop teaching competencies actually prepare persons who possess the repertoire of professional knowledge and skills considered necessary for successful teaching.

Teacher educators have professed concern about the above measurement/evaluation problem but their declarations, for the most part, have been greatly influenced by the assumption that an individual experiencing and completing a program of teacher education would be able to handle children in the classroom with some degree of professional and academic success. There never has been much of anything developed in the way of evidence to support this belief. But in the world of educational accountability, what teacher educators and others have taken for granted about successful teacher behavior and resultant pupil outcomes can no longer be accepted on face value.

---

[1] For readers unfamiliar with "high and low inference observation" the terms are discussed on pages 26-28, 31-32.

[2] JOURNAL OF TEACHER EDUCATION, Vol. 29, No. 2, (March/April 1978) entire issue devoted to Assessing the Competence of Preservice Teachers.

Russ Vlaanderen, TRENDS IN COMPETENCY-BASED TEACHER CERTIFICATION. Denver, Colo: Research and Information Department, Education Commission of the States, Oct. 1980 (unpaged pamphlet)

Sharon G. Boardman and Michael J. Butler (Eds.) COMPETENCY ASSESSMENT IN TEACHER EDUCATION. Washington, D.C.: American Association of Colleges for Teacher Education and ERIC Clearinghouse on Teacher Education, August 1981. 97 pp.

Donald Medley has probably put the matter most succinctly when he stated, "Only when we know why a teacher is effective — as well as how — can we decide how best to train teachers."[3] B. Othanel Smith has indicated that a research knowledge base for pedagogical education exists, and this base can be analyzed and related to the task of preparing better, more skillful teachers. He states: "The research of the last 25 years has now accumulated a body of dependable, generic knowledge with which to begin the development of a more effective program."[4] In another review of the pro's and con's on teacher effectiveness research, Gage has concluded, "We do have some relationships between teacher and pupil achievement and attitudes on which a scientific basis for the art of teaching may be erected."[5] Coker, Medley and Soar, have commented on the power behind the concept of competency-based teacher education and competency-based certification by pointing out that the objectives of a teacher education program for certification procedures can be operationally defined. They state:

> Thus, the teacher (and, indirectly, the teacher education program) can be held accountable, can be required to demonstrate mastery of a specified set of competencies (i.e., teaching behaviors), before being allowed to start teaching.[6]

Leading professionals and researchers involved with teacher education are now making the point that there does exist a valid, limited research base for the development of teacher education programs which can provide those graduating from them with the competencies, (knowledge and skills, plus accompanying attitudes and values) which should be in the repertoire of competent, effective teachers and teaching. What is now needed is additional information and evidence about this emerging research base. The base can only be firmly established and extended through continued research of teacher competence and effectiveness with empirical procedures which go beyond paper-pencil tests or performance ratings. Medley has gone so far to suggest that:

---

[3] Donald M. Medley, TEACHER COMPETENCE AND TEACHER EFFECTIVENESS. A REVIEW OF PROCESS-PRODUCT RESEARCH (Washington, D.C.: American Association of Colleges for Teacher Education, August 1977), p. 6.

[4] B. Othanel Smith and others: A DESIGN FOR A SCHOOL OF PEDAGOGY (Washington, D.C.: U.S. Department of Education, U.S. Government Printing Office, 1980), p.5.

[5] N. L. Gage, THE SCIENTIFIC BASIS OF THE ART OF TEACHING (New York: The Teachers College Press, 1977), p. 30.

[6] Homer Coker, Donald M. Medley and Robert S. Soar. "How Valid Are the Expert Opinions About Effective Teaching?" PHI DELTA KAPPAN, Vol. 62, No. 2 (October 1980), p. 131.

> Close collaboration between the researcher and the teacher educator . . . can produce maximum improvement in the effectiveness of teacher education programs while (instead of after) developing the solid research base that seems so far out of reach today.[7]

Smith's book, *A Design for a School of Pedagogy*, provides teacher education institutions specific knowledge and skills for preservice pedagogical training based on a variety of assembled, classified and recognized research results about teachers, the teaching process and the educational outcomes of such research.[8]

There is no doubt that the problem of assessing the competency of an individual teacher has been neglected and the matter reflects not so much an absence of means to develop valid procedures for assessing teacher competence but a failure to use them. Teacher training institutions not only have the function of training teachers but also a gatekeeping function to perform which requires that they should graduate only those students who demonstrate teaching competence. This second function has been neglected by most teacher training institutions, although they are the only agencies who have the opportunity and the potential resources to make accurate assessments of their graduates' competence before making them available for professional practice. Consequently, better assessment devices and procedures which are more valid, objective and practicable than what has been used in the past are necessary. This monograph purports to show some procedures which can be developed and used, with institutional resources and efforts, to discriminate between competent and incompetent individuals preparing to be teachers.

Figure 1, "A Conceptual Diagram of the Elements Involved in Teacher Research and Evaluation," (taken from Medley with a slight

---

[7] Ibid p. 70.
[8] Op Cit p. 49-106.

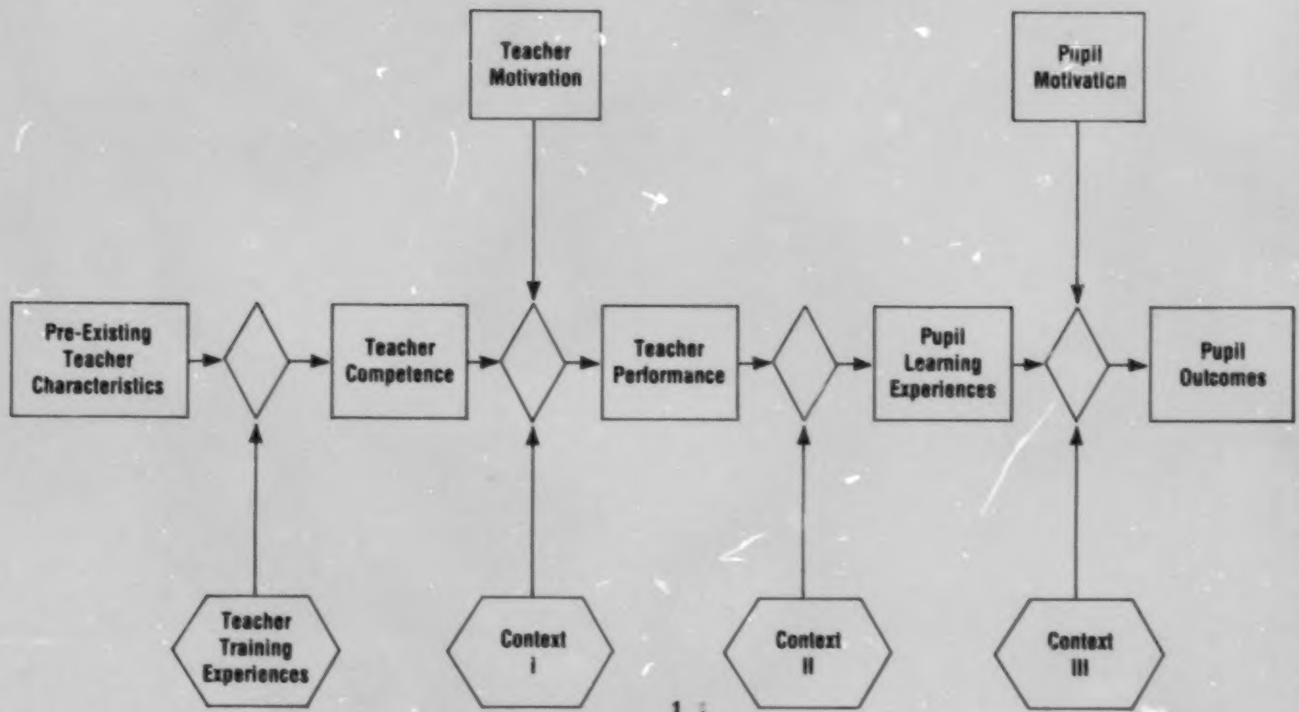

Teacher Motivation
Pupil Motivation
Pre-Existing Teacher Characteristics
Teacher Competence
Teacher Performance
Pupil Learning Experiences
Pupil Outcomes
Teacher Training Experiences
Context I
Context II
Context III

Figure 1. A Conceptual Diagram Of The Elements Involved In Teacher Education Research And Evaluation*

*Adapted from Medley, see footnote 9

modification in name[1]) illustrates the basic differences between the terms "Teacher Competence," "Teacher Performance," and "Teacher Effectiveness." Figure 1 indicates that there are five points in teacher preparation and service where teacher assessment can occur to provide information on teaching competence, performance and effectiveness. Teachers can be assessed at any of these five points, but the point used will depend upon the purpose for which the assessment is made.

Before discussing these assessment points, brief definitions of teacher competence, performance and effectiveness are in order. "Teacher competence" is defined in terms of the repertoire of professional knowledges, professional skills and professional value positions the teacher brings to the job. "Teacher performance" is defined in terms of the teaching behavior of the teacher on the job. "Teacher effectiveness" is defined in terms of the impact the teacher has on the pupils the teacher teaches. In making these distinctions, how well a teacher performs and how effective the teacher really is depends upon a variety of situational or contextual factors as well as teacher competency and motivation. Having made these definitions, let us return to consideration of the assessment points.

The first assessment point concerns the characteristics that an individual brings to teacher preparation and these can be measured when the student first declares that he/she wishes to enter the teaching profession. After initial entry into the profession, some of these characteristics will persist unchanged and others will be modified by the experiences the individual has in teacher training and inservice teacher activities.

The second assessment point occurs at the end of the teacher training experience in what is known as the student teaching situation or experience. Teacher competence can be directly assessed at the culminating experience (student teaching) of any teacher training program.

The third assessment point is an examination of teacher behavior (teacher performance) on the job in the first year of actual inservice teaching. The influences which can affect such teacher performance are the pre-existing teacher characteristics, previous teacher training experiences, and two additional factors which have heretofore not been involved — the situational context in which the teacher works and how the teacher feels about the job, that is teacher attitudes/motivation.

---

[1] Donald M. Medley, "The Role of Teacher Educators in Teacher Competency Assessment," INSERVICE, (Newsletter of the National Council of States on Inservice Education) Syracuse, N.Y.: School of Education, Syracuse University, Sept. 1981, p. 11.

The fourth and fifth assessment points concern the measurement of teacher effectiveness or the evaluation of teachers not on the basis of their own behaviors but rather on the learning experiences and educational outcomes of their pupils. Specifically, the fourth assessment point, labeled "Pupil Learning Experiences" in Figure 1, assumes that an assessment of such experiences is affected, directly or indirectly, by a variety of influences, only one of which is teacher performance. Such influences are pupil motivation and various contextual variables intrinsic to or impinging upon the learning situation. The fifth and final assessment situation is the measurement of pupil knowledge, abilities and/or attitudes after instruction. Again, some of these outcomes are attributable to the teacher but other variables are also present including a third type of context variable which interacts with pupil learning experiences to determine how much a pupil will learn from such experiences. It is impossible to identify or analyze all such variables or the way in which they all might be controlled in school settings. Again, Medley puts the matter properly when he states:

> "The major point to be learned (from Figure 1) about the contextual variables is that more of them affect evaluations of teacher effectiveness than affect evaluations of teacher performance, and as a result, the valid measures of either of them are much more difficult to obtain than valid measurements of teacher competence."[10]

Teacher competence can be assessed most directly at the second of the five basic boxes in the flow chart portrayed in Figure 1. Teacher performance is best assessed at box 3, while boxes 4 and 5 represent indirect assessments of competence based on the assumption that the closer one gets to pupil outcomes, the more valid becomes the evaluation of teacher competence. This may not be the case when all contextual variables receive proper consideration. At this stage in the measurement of teacher competence and effectiveness, we are most skilled and comfortable with the assessment of teacher competence (box 2), as that competence is demonstrated by prospective graduates of a teacher education program in their student teaching experiences.

---

[10] Donald M. Medley, "The Role of Teacher Educators in Teacher Competency Assessment," COMPETENCY ASSESSMENT IN TEACHER EDUCATION: MAKING IT WORK, CONFERENCE RESOURCE BOOK. (Complied by Donna Gollnick, Sharon McMullin and Karl Massanari). (Washington, D.C.: American Association of Colleges for Teacher Education 1980), p. 9.

The foregoing concepts and distinctions are important in any consideration of measuring teacher competence. In the present era of state legislation or other public demands requiring assessment of the "competence of teachers" (preservice and inservice) for certification or any purpose, the distinctions between competence, performance and effectiveness, as indicated by Medley, are usually fuzzy and confused on the part of those attempting assessment procedures. When one tries to infer teacher competency on the basis of pupil learning experiences and outcomes, there are considerable problems in any such measurement effort. There are other multiple factors not controlled by the teacher which affect pupil outcomes directly or indirectly. These must be taken into account in any measurement attempt involving what pupils learn and the relationship of such information to teacher competence. In the present state of the art, it is likely not possible to take all such variables into consideration because not enough is known about how to measure all of them, and it would be too costly to do so even if such knowledge was available. Consequently, it seems best to concentrate on the "teacher competence" that *can* be measured and *where* it can be measured. And since teacher competence is fundamentally defined in terms of what a teacher knows and what can be behaviorally demonstrated on the basis of that knowledge, the best procedure to research and evalute teacher competence would seem to be through observing what a teacher does, on the basis of what they have been taught in their teacher education program, rather than inferring an effect of such performance by observing pupil behavior and/or measuring pupil outcomes. This is what is occurring with the student teachers of the competency-based teacher education program at The University of Toledo and what will ultimately be followed up and similarly measured in these students' initial teaching positions after graduation from the CBTE program.

Chapter II

A LIMITED REVIEW OF RELEVANT LITERATURE

The General Concern for Competent, Effective Teachers

The study of teachers, teaching effectiveness and the educational programs to produce them has and continues to be a constant concern. Much of the attention to this concern has been provided in broad surveys, reports, or opinions by individuals or organizations which, it was hoped, would result in improved programs for educating teachers and, consequently, better practitioners in the classroom. In the past 40 years, these efforts have been joined by an increasing volume of research effort to look at teacher characteristics, teacher behavior, and the relationship of these phenomena to improved learning on the part of children in classroom learning situations.

The national surveys and reports about teacher education have attempted to stimulate teacher education improvement by basically looking at the political, social, economic and various external/internal forces having an impact on pedagogical education. These reports beginning with the *Commonwealth Teacher-Training Study* published in 1929[11] through the report of the *Bicentennial Commission on Education for the Profession of Teaching*[12] have all tended to reflect the conflicts and tensions of the time in which the reports were written. The Commonwealth Teacher-Training Study analyzed more than 2000 statements of teacher activities and came to the conclusion that six major characteristics about teachers were most important: good judgment, self control, considerateness, enthusiasm, magnetism and adaptability.[13] In contrast, the Bicentennial Commission report (which culminated two years of study) was concerned with teacher preparation, an emphasis on a career-long pursuit of teacher effectiveness, the status of teaching as a profession, the influences which control teacher education and the control of professional quality. In between there were various, similar pronouncements. For example, in the middle

---

[11] W.W. Charters and Douglas Waples, COMMONWEALTH TEACHERS TRAINING STUDY (Chicago: University of Chicago Press, 1929).

[12] Robert B. Howsam, Dean C. Corrigan, George W. Denemark and Robert J. Nash, EDUCATING A PROFESSION (Washington: American Association of Colleges for Teaching Education, 1976).

[13] W.W. Charters and Douglas Waples, OP CIT., p. 18.

40's, the American Council on Education, Commission on Teacher Education urged (1) making the democratic way of life a basic orientation for teacher education consideration and (2) requiring that teachers be more sensitive to social realities and ideals in respect to schooling.[14] The Conant study backed by the Carnegie Corporation made very specific recommendations about certification, academic preparation of teachers, theory and practice of teaching, course requirements and the personnel needed for such activities.[15] The National Commission on Teacher Education and Professional Standards Task Force which produced *New Horizons for the Teaching Profession* provided direction on professional standards, pre and inservice teacher education, accreditation, certification and admission and retention policies.[16] The report titled *Teachers for the Real World* put emphasis on cultural and ethnic factors in teacher education as well as pedagogical concerns.[17]

All such reports and others like them addressed improved teacher education programs incorporating a variety of pedagogical emphases and innovations which mirrored American educational concerns, but none of them provided much of anything in the way of fundamental information about process-product research, the probable cause-and-effect relationship between teacher behavior and pupil behavior and learning. This type of research was left to isolated studies. In the past ten years there have been significant compilations of these which provide more useful information than that generated through general reports and pronouncements on the subjects.

A General Review of Research on Teacher Competency, Performance and Effectiveness

The history of research on teacher effectiveness provides indications of researchers' conceptions of the nature of effectiveness and why the early studies in this area were not particularly useful and generally led to inconclusive results. Teacher effectiveness was first perceived in research as a consequence of characteristics or personality traits of teachers and attempts were made to identify those characteristics. A second direction of teacher effectiveness research involved the methods of teaching being used. A third aspect of such inquiry concerned the maintenance of classroom climate and conditions. Most recently, teacher effectiveness research has been involved with teacher competencies and the mastery of them. Consideration is now being given to the ability of teachers to demonstrate teaching competencies and the possible resultant effects of such performance on pupil behavior and learning.

---

[14] Commission on Teacher Education, TEACHERS FOR OUR TIMES. (American Council on Education, 1944). Commission on Teacher Education, THE IMPROVEMENT OF TEACHER EDUCATION. (American Council on Education, 1946).

[15] James Bryant Conant, THE EDUCATION OF AMERICAN TEACHERS. (New York: McGraw-Hill Book Company, 1963).

[16] Margaret Lindsay (Ed.), NEW HORIZONS FOR THE TEACHING PROFESSION. (Washington, D.C.: National Commission on Teacher Education and Professional Standards, National Education Association, 1961).

[17] B. Othanel Smith and others, TEACHERS FOR THE REAL WORLD. (Washington, D.C.: American Association of Colleges for Teacher Education, 1969).

Research on teacher characteristics was culminated by the now classic study of Ryans' for the American Council on Education titled, *Characteristics of Teachers.*[18] This study, widely reviewed and quoted, indicated patterns of teacher behavior, attitudes and qualities (on the basis of classroom observation) which characterized elementary and secondary teachers. Although the study was not directed specifically to the problem of teacher effectiveness (as the term is now defined), the matter was not entirely ignored as the characteristics of teachers were classified as high, average, or low in respect to certain types of classroom behavior. Nevertheless, there was no evidence to show that teachers possessing particular characteristics were actually more or less effective in helping pupils achieve various educational goals.[19]

Research on so-called effective methods of teaching (methods experiments) proved to be as inconclusive or contradictory as research on teacher characteristics. Generalizations about effective teachers were difficult to make from methods experiments when the unit of analysis was not focused on the teacher but rather on the mean gains in test performance of classes taught by different teachers using different methods. Efforts were not made in this research approach to determine the consistency of class results by different teachers using the same method.

In 1964, Cyphert and Spaights concluded an extensive research project that analyzed research in teacher education conducted over the prior ten-year period. They found that research in teacher education tended to focus on five main areas, namely:

---

[18] David G. Ryans, CHARACTERISTICS OF TEACHERS. (Washington, D.C.: American Council on Education, 1960).

[19] J. W. Getzels and T. W. Jackson, "The Teacher Personality and Characteristics," in N. L. Gage (Ed.), HANDBOOK OF RESEARCH ON TEACHING. (Chicago: Rand McNally and Company, 1963, p. 574.

1. Present practices in teacher education.
2. Characteristics of personality associated with various teacher behaviors.
3. Aspects of training and experiences associated with different kinds of teaching behavior or criteria of teaching success.
4. Comparative effects of different methods of educating teachers and other educational personnel.
5. The emergence of the behavioral systems in which methods were developed for recording and analyzing effective teacher performance to the end that this information can serve as a decision-making base for selecting desireable teacher education content.[20]

A research direction in teacher effectiveness was emerging in the middle 1960's which focused on both teacher behavior and pupil learning — an effort which became known as process-product research. Such thinking had occurred much earlier than the 1960's. Barr and his associates had advocated the use of criteria of teaching effectiveness based on measured pupil gains rather than expert opinion.[21] The work of Flanders and Amidon in "Interaction Analysis" provided a much utilized process for recording teacher behavior.[22]

The first *Handbook of Research on Teaching*[23] contained several chapters about process-product research studies. In 1971, Rosenshine's review of 50 such studies provided evidence that some aspects of teaching style were related to pupil learning.[24]

---

[20] Frederick R. Cyphert and Ernest Spaights, AN ANALYSIS AND PROJECTION OF RESEARCH IN TEACHER EDUCATION, USOE Cooperative Research Project No. F-150. (Columbus, Ohio: The Ohio State University Research Foundation, 1964), pp. 105-106 and 303.

[21] Avril S. Barr and Lester M. Emans, "What Qualities are Prerequisites to Success in Teaching?" NATION'S SCHOOLS, Vol. 6, (September 1930) pp. 60-64.

[22] Ned A. Flanders, TEACHER INFLUENCE, PUPIL ATTITUDES AND ACHIEVEMENT, Final Report, Cooperative Research Program, Project #397. (Minneapolis, Minn.: University of Minnesota, 1960).

[23] N. L. Gage (Ed.), HANDBOOK OF RESEARCH ON TEACHING. (Chicago: Rand-McNally Company, 1963).

Since the early 70's there have been published a large number of studies concerned with process-product research and the broad concern for teacher effectiveness. A complete review and analysis will not be provided. The majority have been summarized and reviewed in various publications, some principal and useful sources being

> S. N. Bennett, "Recent Research on Teaching: A Dream, A Belief, and a Model," *British Journal of Educational Psychology*, Vol. 48, Part 2, (June 1978), pp. 127-147.
>
> Gary D. Borich and Kathleen S. Fenton, *The Appraisal of Teaching: Concepts and Process*. (Reading, Mass.: Addison-Wesley Publishing Co., 1977).
>
> J. E. Brophy, "Teacher Behavior and its Effects," *Journal of Educational Psychology*, Vol. 71, No. 6, (December 1979), pp. 733-350.
>
> N. L. Gage, *The Scientific Basis of the Art of Teaching*. (New York: Teachers College Press, Teachers College, Columbia University, 1978).
>
> Thomas L. Good, "Teacher Effectiveness in the Elementary School," *Journal of Teacher Education*, Vol. 30, No. 2, (March-April 1979), pp.52-63.
>
> Donald M. Medley, *Teacher Competence and Teacher Effectiveness*. (Washington, D.C.: American Association of Colleges for Teacher Education, August 1977).
>
> Penelope L. Peterson and Herbert J. Walberg, (Eds.), *Research on Teaching*. (Berkeley, Calif.: McCutchan Publishing Corporation, 1979).
>
> Thomas F. Ryan, *Extending Teacher Education Programs: A Five Segment Continuous Development Program*

---

[24] Barak Rosenshine, TEACHING BEHAVIORS AND STUDENT ACHIEVEMENT. (Slough, England: National Foundation for Educational Research in England and Wales, 1971).

for Teachers. Unpublished manuscript of a presentation to the Teacher Education Council of State Colleges and Universities, Louisville, Kentucky, October 30, 1980, Appendix A, pp. A-1-A10.

B. Othanel Smith, *A Design for a School of Pedagogy*. (Washington, D.C.: United States Department of Education, U.S. Government Printing Office, 1980), pp. 79-99.

Robert M. Travers, (Ed.), *Second Handbook of Research on Teaching*. (Chicago: Rand-McNally and Company, 1973), Chapters 5, 6 and 7.

Ian Westbury, "Research into Classroom Processes: A Review of Ten Years Work," *Journal of Curriculum Studies*, Vol. 10, No. 4, (October-December 1978), pp. 283-308.

A number of the above are of particular interest to teacher education because they have taken the present process-product research knowledge and have translated this information into a series of statements about definite relationships found between teacher behavior and its effects on pupil behavior and achievement. From such findings, better teacher education programs can be developed and current programs modified. Two sources (Medley and Smith) are paticularly useful in this respect. These and a few others will receive comment along with pertinent research findings.

Medley's study on *Teacher Competence and Teacher Effectiveness*, published by the American Association of Colleges for Teacher Education, is a useful compilation of process/product relationships. From an original list of 732 studies, a weeding-out process provided 289 which purported to provide information on the question, "How does the behavior of effective teachers differ from that of ineffective teachers?" Four demanding criteria were then used to determine which of the 289 would remain for final review. Among the criteria were (1) generalizability, (2) measures of long-term pupil gain, (3) the behaviors exhibited could be reproduced, and (4) the relationship had to be both reliable enough to be statistically significant and large enough to be practically significant. For a behavior to have an "important" relationship to teacher effectiveness it was arbitrarily decided that the minimum overlap in the variances of the two measures involved would not be less than 15 percent which corresponds to a correlation coefficient of .39.

In addition, no relationship was considered, whatever its magnitude, unless it was statistically significant at the .05 level. The application of the criteria resulted in the selection and thorough review of 14 studies. It is interesting to note that the procedures applied resulted in including the four major studies on teacher effectiveness (Brophy and Evertson, 1974,[25] McDonald and Elias, 1976[26] Soar, 1973,[27] and Stallings and Kaskowitz, 1974[28]) which most major researchers have agreed were the most prominent in the field. Gage, for example, has stated that these four studies "...may be regarded as providing a capstone of the correlational approach to meeting the need for an improved scientific basis for the art of teaching and the education of teachers."[29]

The studies analyzed by Medley were necessarily limited by the fact that most of them dealt with the primary grades and the teaching of reading or arithmetic. However, certain generalizations emerged about teacher performance and effectiveness.

The learning environment in an effective teacher's classroom was what might be expected — it tended to be orderly, psychologically supportive and maintained with a minimum of teacher effort. There was less deviant, destructive behavior, fewer teacher rebukes and less criticism in effective teachers' classes. These teachers used more praise, reinforcement and encouragement than did ineffective teachers. The effective teacher was the better classroom manager, less permissive, more supportive and maintained a learning environment freer from disruptive pupil behavior. The maintenance of a good learning environment provided some inferences about teacher planning, lesson organization and classroom operation.

---

[25] Jere E. Brophy and Carolyn M. Evertson, PROCESS-PRODUCT CORRELATIONS IN THE TEXAS TEACHER EFFECTIVENESS STUDY: FINAL REPORT. Research Report No. 74-4. (Austin, Texas: Research and Development Center for Teacher Education, University of Texas, 1974).

[26] Frederick J. McDonald and Patricia Elias, THE EFFECTS OF TEACHING PERFORMANCE ON PUPIL LEARNING, BEGINNING TEACHER EVALUATION STUDY: PHASE II, FINAL REPORT, VOL. 1. (Princeton, N.J.: Educational Testing Service, 1976).

[27] Robert S. Soar, Final Report, FOLLOW THROUGH CLASSROOM PROCESS MEASUREMENT AND PUPIL GROWTH (1970-1971). (Gainesville, Fla.: Institute for the Development of Human Resources, University of Florida, 1973).

[28] Jane Stallings and D. Kaskowitz, FOLLOW THROUGH CLASSROOM OBSERVATION EVALUATION, 1972-1973, SRI Project, URU-7370. (Menlo Park, Calif.: Stanford Research Institute, 1974).

[29] N. L. Gage, "An Experiment on Teacher Effectiveness and Parent-Assisted Instruction in the Third Grade: Background and Rationale," in AN EXPERIMENT ON TEACHER EFFECTIVENESSS AND PARENT-ASSISTED INSTRUCTION IN THE THIRD GRADE, PROGRAM ON TEACHING EFFECTIVENESS. (Stanford, Calif.: Center for Educational Research at Stanford, School of Education, Stanford University, May 1978). p. 1.

The use of pupil time (time-on-task) was an important factor in teacher effectiveness. The ineffective teacher spent the least time on "academic" activities, had the most small-group activity and engaged the pupils in considerable independent study. The effective teacher's classroom utilized more large-group activity, devoted more time to academic concerns and had less independent seat work. However, what seat work was engaged in received closer supervision in the effective teacher's classroom than that which occurred in an ineffective situation.

Teachers who asked the most high-level and fewest low-level questions, had the most pupil questions, provided the most feedback to pupil questions and who amplified or discussed pupil-initiated statements were not found to be the most effective teachers. Effective teachers asked more low-level questions, initiated fewer questions from pupils, provided them less feedback and did not tend to expand on what pupils said.

An interesting finding was that apparently successful teacher behaviors differed markedly for higher and lower socioeconomic status (SES) pupil groups. What turned out to be effective teacher behavior with one group proved ineffective with the other. Another interesting SES inference was the matter of attention given to individual pupils. The teacher who gave the most individual attention to pupils in high SES classes became the least effective in producing cognitive gains; the reverse was apparent in low SES classes. The effective teachers of low SES pupils in the primary grades engaged their children in different discussion and questioning strategies than those used by effective teachers with high SES pupil groups. It seemed that pupils from different SES backgrounds demanded different teaching strategies to maximize learning activities.

Some of these research results may appear surprising because they challenge long-held assumptions and beliefs about effective teaching. When sufficient results have been obtained on the subject, teacher educators will be required to incorporate them into their program development and operation.

The volume by Smith, *A Design for a School of Pedagogy*, is fundamentally a reasoned statement that teacher education and pedagogical practice must be based more than they have been on research findings about teacher effectiveness. Smith states, "As a profession matures and its research base becomes more and more highly developed, more and more knowledge tends to come from researchers and less from practitioners."[30] He further emphasizes, "...the need to assemble what we know from research, process-product studies as well as other types, and how to relate it to program development, policymaking, and pedagogical training."[31] Chapter 10 of Smith's volume provides a compilation of basic generic knowledge about teaching gleaned from process-product research which has definite implications for pedagogical training. Under the general labels of diagnosis, planning, management of time, space and resources, management of instruction, management of student conduct, communication and evaluation, Smith provides fundamental research information on these topics,[32] which contain 18 useful research findings about effective teaching.

The volume by Borich and Fenton, *The Appraisal of Teaching: Concepts and Process*, contains an excellent list of selected readings by prominent researchers or reviewers of research on teacher competencies, performance and performance appraisal systems. Chapter Six provides the reader accounts of the five large-scale teacher effectiveness studies by Brophy and Evertson, Soar and Soar, Stallings and Kaskowitz, Good and Grouws and McDonald and Elias. Chapter Seven on measuring teacher performance provides considerable review and criticism of present measurement efforts by such authors as Berliner, Soar, Popham and Scriven.

Good's article on "Teacher Effectiveness in the Elementary School," summarizes some conclusions from process-product studies. From his review of the research he states: "...(1) elementary school teachers do not exert differential effects upon student achievement, (2) classroom management skills are exceedingly important and relate positively to student achievement in every process-product study conducted to date and (3) a pattern of teaching behavior called direct instruction appears to be a useful heuristic for describing effective elementary classroom teachers."[33]

---

[30] B. Othanel Smith, OP.CIT., p. 50.
[31] IBID. p. 51.
[32] IBID. pp. 93-94.

In a presentation to the Teacher Education Council of State Colleges and Universities in Louisville, Kentucky, on October 30, 1980, Thomas F. Ryan provided a review of research on teacher affect which was excerpted from an unpublished manuscript developed by George Denemark and Norma Mutter. From this source and others (as indicated), a number of major findings and/or conclusions from research on teacher effectiveness are listed below.

1. Real effects may be expected from differences in teacher behaviors and other instructional variables, but the effects that may be expected (and detected) are probably more modest than formerly thought. (Ryan, from Stallings and Kaskowitz)
2. There apparently is no single effective instruction method or individual teacher behavior. Teaching is a complex act and must be examined as constellations of behavior and as behavior occuring in a larger context. (Ryan, from McDonald and Elias, Brophy and Evertson, Gage and Good)
3. Effective teacher behavior varies according to the subject being taught. The amount of positive teacher influence possible may vary according to subject area. (Ryan, from McDonald and Elias and Soar)
4. Effective teacher behavior varies according to pupil characteristics such as age and socioeconomic status. The impact of teacher behaviors on learning may be greater for some types of students than for others. However, Gage is reported as concluding from research findings that there is support for both generality and specificity. (Ryan, from Brophy and Evertson, McDonald and Elias, Medley and Gage)
5. The stability of measures of teacher effectiveness and teacher behaviors is low to moderate from year to year but teachers who have low effectiveness tend to remain

---

[33] Thomas L. Good, "Teacher Effectiveness in the Elementary School," JOURNAL OF TEACHER EDUCATION, Vol. 30, No. 2 (March-April 1979). p. 53.

low in effectiveness. (Ryan, from Brophy, Berliner and Good)

6. Individual teachers do make a difference; differences among teachers are substantial, and relatively effective and ineffective teachers can be identified for study. (Ryan, from Brophy and Evertson, Berliner and Tikunoff and McDonald and Elias)

7. "Classroom management" can be organized into principles by which the consequences of behaviors can be predicted and salient missing behaviors identified and developed. (Ryan, from Brophy and Evertson and Good)

8. Teachers who direct classroom activities without giving students choices, approach the subject matter in a direct, business-like way, organize learning around questions they pose and occupy the center of attention, will generally have students who show greater task orientation and greater gains and achievement, creativity, inquiry, writing ability and self-esteem. (Smith)

9. When teachers adjust the cognitive demands of their lessons to the level of students' capabilities, student achievement increases. (Smith)

10. When teachers make use of probing techniques in instruction, students will make more inferences and more accurate inferences concerning subject matter. (Smith)

11. Teachers who give complete instructions and demonstrations of new materials and assignments and check to see that the students do understand how to do the assignments before releasing them to work on their own generally have students who will show lesson gains and subject matter achievement. (Smith)

12. Direct instruction, increasingly being supported by research, requires the setting and articulating of learning goals, actively assessing student process, and frequently making class presentations illustrating how to do the assigned work. A major dimension of directing

instruction seems to be time, with the crucial time for gains in student achievement being that in which students are actively engaged in instructional activity. (Ryan, from Good and others)

13. Classroom climate and affective behaviors are important in teaching. Open approaches have often been found superior in promoting positive attitudes toward school and teacher, independence, creativity and curiosity. (Ryan, from Tikunoff, Berliner and Rist, Stallings and Kaskowitz and Soar)

14. Teacher expectations of students apparently influence teacher behavior and student achievement. Inappropriate expectations or attitudes can cause teachers to behave in self-defeating ways with negative teacher behaviors being exhibited toward low expectation students. Less is expected and less is taught. (Ryan, from Brophy, Brophy and Evertson and McDonald and Elias)

No account of research on teacher effectiveness would be complete without mention of the findings of the Beginning Teacher Evaluation Study (BTES) conducted for the California Commission on Teacher Preparation and Licensing. This study which spanned six years concluded in 1978. The study consisted of two parts with the first research effort and report being provided by McDonald and Elias[34] and the second study conducted by a Far West Laboratory research team.[35] The major findings of the study were as follows under the headings of Academic Learning Time and Student Achievement (Items 1-5) and Instructional Processes and Classroom Environment (Items 6-14).

" 1. The amount of time that teachers allocate to instruction in a particular curriculum content area is positively associated with learning in that content area.

" 2. The proporation of allocated time that students are engaged is positively associated with learning.

---

[34] Frederick J. McDonald and Patricia Elias, LOC CIT.

[35] C. W. Fisher, D. C. Berliner, N. N. Filby, R. Marliave, L. S. Cahen, M. M. Disaw and J. E. Moore, TEACHING AND LEARNING IN THE ELEMENTARY SCHOOLS: A SUMMARY OF THE BEGINNING TEACHER EVALUATION STUDY. (San Francisco: Far West Laboratory for Educational Research and Development, September 1978).

“ 3. The proportion of time that reading or mathematics tasks provide a high success rate for a student is positively associated with student learning.

“ 4. The proportion of time that reading or mathematics tasks provide a low success rate for a student is negatively associated with student learning.

“ 5. Increases in Academic Learning Time are *not* associated with decreases in attitude toward mathematics, attitude toward reading, or attitude toward school.

“ 6. The teacher’s ability to diagnose student skill levels is related to student achievement and Academic Learning Time.

“ 7. The teacher’s ability to prescribe appropriate tasks is related to student achievement and student success rate.

“ 8. More substantive interaction between the student and an instructor is associated with higher levels of engagement.

“ 9. Academic feedback is positively associated with student learning.

“10. Structuring of the lesson and giving directions on task procedures were positively associated with student success rate.

“11. Explanation specifically in response to student need is negatively associated with student success rate.

“12. More frequent reprimands for inappropriate behavior are negatively associated with student learning.

“13. The teacher’s value system is related to Academic Learning Time and to student achievement. Teacher emphasis on academic goals is positively associated with student learning.

> "14. A learning environment characterized by student responsibility for academic work and by cooperation on academic tasks is associated with higher achievement."[36]

While research on teacher effectiveness does hold promise for teacher education there are also limitations in such research. Most of such research has been limited to basic skills developed in reading and mathematics in the elementary grades. Present findings are not based, for the most part, on research performed in grades beyond the elementary school. Work is still continuing to define and clarify the variables in determining teacher effectiveness. Pupil achievement, for the most part, has been expressed as scores on standardized achievement tests, with such scores correlating highly with scores on standardized intelligence tests. Achievement so measured is considered questionable when used with very young and/or culturally different children.

Research on teacher competence, teacher performance, and teacher effectiveness has shown considerable productiveness in the decade of the 70's. Researchers are now concluding that at least a beginning data base for the content of teacher educaton exists, with results substantial enough to provide direction for teacher education efforts. The writings of Bennett, Gage, Good, Medley and Smith which have been partially reviewed and footnoted all express this viewpoint. Stallings has perhaps expressed this point of view best in a recent statement:

> "The research on teaching has been driven not so much by theory as a curiosity about the nature of effective and ineffective teaching process. In all cases, the researchers went to classrooms and observed what teachers and students were doing. They studied what was occurring in classrooms where students were making gains, as well as where students were not making gains. Some of the studies used qualitative data, others used quantitative data, some were conducted concurrently, other sequentially. The communication among those who conduct research on teaching tends

---

[36] Carolyn H. Denham (Ed.), "The Major Findings of the Beginning Teacher Evaluation Study," THE GENERATOR OF DIVISION G. The Social Context of Education, the American Educational Research Association, Vol. 9, No. 1, (Spring 1979), pp. 5-6.

> to be frequent and constructive. As new studies are mapped, methodologies and findings are compared to explore new frontiers of the teaching and learning territory. Thus, a body of knowledge regarding the management of time, tasks and students is accumulating and can be useful in guiding the training of inservice and preservice teachers."[37]

The ongoing research on teacher behavior and competence at The University of Toledo is another example of an attempt to add to the body of knowledge on the teaching process of concern to Stallings and similar persons. The UT researchers hope their efforts make a significant contribution to the "new frontiers of the teaching and learning territory."

---

[37] Jane Stallings, "Allocated Academic Learning Time Revisited, or Beyond Time-on-Task," EDUCATIONAL RESEARCHER, Vol. 9, No. 11, (December 1980), p. 16.

Chapter III

THE IDENTIFICATION OF TEACHING COMPETENCIES AND THE METHODOLOGY OF CLASSROOM OBSERVATION

The Identification of Teaching Competencies

A basic question when conducting research on teacher effectiveness is, "What are the behaviors that should be measured by the observation?" In the context of a competency-based teacher education program the analogous question is, "What are the competencies which underlie the program?" In a more general sense the question could be phrased, "What competencies or behaviors do we want demonstrated by teachers or prospective teachers?" It is the matter of knowing what to measure before measurement is attempted.

The College of Education and Allied Professions of The University of Toledo designed a competency-based teacher education (CBTE) program for the preparation of elementary teachers in 1968-69 and began the implementation of that program in 1970. At the same time (1970-72), a professional program for the preparation of secondary teachers in the CBTE mode was designed and its implementation begun. By 1973, the College was operating a fully developed CBTE program which had basically evolved from a research effort funded by the United States Office of Education. A complete description of the developmental effort appears in the volume *Partners for Educational Reform and Renewal.*[38]

Description of The University of Toledo CBTE Program

The competency-based portion of The University of Toledo teacher education program consisted of all professional course work in the program comprising three "career decisions" pre-professional courses taken by both elementary and secondary teacher education students, four "block" courses for elementary trainees followed by all day, full-quarter student teaching, and two "block" courses for secondary trainees followed by all day, full-quarter student teaching. All of these courses were accompanied by clinical, field-based experiences with children in schools, amounting to a total of 600 hours of clinical, field-based experience. The course work in the courses was organized in modular form and each course was taught by a faculty team consisting of four or more faculty members assisted by graduate assistants. A "block" course usually met two and one-half to three hours per day four days a week. The modules in the courses and student teaching experiences contained the competencies the students were expected to achieve and demonstrate. These competencies were expressed as behavioral objectives for the various instructional modules. A procedure to identify the competencies, as reflected in the behavioral objectives, was devised and implemented.

---

[38] George E. Dickson and Richard W. Saxe, PARTNERS FOR EDUCATIONAL REFORM AND RENEWAL. (Berkeley, Calif.: McCutchan Publishing Corp., 1973).

Competencies Based on Content Analysis

In the context of The University of Toledo's CBTE program, the focus was on measuring the competencies underlying the program. The basic approach to identifying program competencies was to make a thorough content analysis of the course modules. The analysis was conducted by three faculty members familiar with the CBTE program. Two faculty members were specialists, respectively, in elementary education and secondary education, and the third was the director of the original project to design the CBTE program at The University of Toledo. All instructional modules of the basic CBTE program were reviewed. These modules comprised the content of 11 basic courses; three at the secondary level, five at the elementary level and three courses common to both levels.

CBTE program analysis produced 96 competencies. These were ultimately reduced to a basic list of 49 competencies. The 49 competencies identified are found in Appendix A. This synthesis was accomplished by the elimination of duplications, the combining, condensing, refining and reformulating of similar competency statements, and the process of mapping the competencies into the behaviors measured by the instruments used in this research effort. The final section of this chapter provides additional information on the mapping procedure.

The above procedure involved identifying competencies as they were reflected in the behavioral objectives of the modules. However, it was felt that a broader base of consensus was necessary other than the efforts of three faculty members. Concurrent with the competency identification process involving the CBTE program at The University of Toledo, another University of Toledo faculty team was involved with a statewide study[39] to identify generic competencies of effective teaching as related to the *Standards for Colleges or Universities Preparing Teachers*[40] operating in the State of Ohio. In this study, a large list of teacher competencies had been identified in relation to the *Standards*. The criterion for including a competency was that it was essential for enhancing learning. However, the list was in an intermediate stage and required further synthesis to eliminate duplications and to reformulate similar competency statements. The result of this content analysis process was a list of 203 competency statements under the following general headings:

Instruction and Managing Instruction
Use of Instructional Materials
Improving the Instructional Program
Improving the Learning Environment
Being a Professional
Planning Activities Prior to Instruction
Evaluation

The statewide list of 203 items was further reduced to 79 competencies by a seven member committee of Ohio educators. A statewide survey was then conducted of over 1,800 Ohio educators, randomly selected with 77 percent being teachers, in which respondents were again asked to rate each of the 79 competencies as to its being essential for enhancing learning. Of the 79 competencies, 37 were rated high enough (2.40 or greater on a 3.0 scale) to be considered essential. The 37 Ohio generic competencies are found in Appendix B.

The statewide study was important in providing a broader source of competency statements. More importantly, when the 37 statewide competencies were compared with the final University of Toledo list of 49 a great deal of similarity was apparent. In essence, the results of the statewide survey and the comparison of these results with The University of Toledo competencies provided additional evidence of the "content validity" of The University of Toledo competencies. The University of Toledo competencies were consistent with those identified in the

---

[39] Thomas Gibney, et al, THE IDENTIFICATION OF GENERIC CONCEPTS OF EFFECTIVE TEACHING AS RELATED TO THE OHIO STANDARDS, Final Report. (Toledo, Ohio: The University of Toledo, June 1980).

[40] STANDARDS FOR COLLEGES OR UNIVERSITIES PREPARING TEACHERS. (Columbus, Ohio: Ohio Department of Education, 1975).

State *Standards* which are required in all Ohio teacher education programs. Although the process was quite subjective, initial validation was apparent.

The Methodology of Classroom Observation

The recording of data from classroom observation has taken various forms, running the continuum from tabulations in matrices representing classroom interaction to recording anecdotal data as subjectively perceived by the observer. More recently, the terminology has become low-inference and high-inference, as characterized by the amount of inference required from the observers. Some writers would argue that the extent of inference is on a continuum, and there may be some merit to this argument. For the purposes of this research, we will maintain the low-inference, high-inference dichotomy, designating any observation that requires the observer to make a qualitative judgment of the classroom interaction as high-inference observation. Observation requiring only a tabulation of defined behaviors is considered low-inference observation.

The advocates of low-inference observation and those for high-inference observation provide convincing arguments for their respective approaches. Advantages cited by one are often viewed as disadvantages by the other. Undoubtedly both approaches have their merits and weaknesses. Presently, indications are that the jury is still out as to the better approach. The method of observation is a very important matter, however. In the final analysis, it is the way measurement is conducted that operationally defines whatever is measured by the observation.

Low-Inference Observation

Low-inference observation generally involves recording the occurrence of specific behaviors which are defined with sufficient precision so that observers will agree on whether or not they have occurred. It is a behavioristic approach throughout. To know what occurs in the classroom, it is necessary to observe the teacher and students. Low-inference observation does not deal with nonclassroom, supportive activities such as writing lesson plans. The assumption is that such activities by themselves do not impact directly on student learning. In order for a behavior to have an effect it must be demonstrated in the instructional context.

Low-inference observation has an advantage in its objectivity. There may be some subjectivity involved in placing an observed behavior in a matrix, but generally this is relatively minor when using trained observers, as must be the case with low-inference observation. There is no option for putting a qualitative judgment on the behavior.

Low-inference observation is generally more straightforward than high-inference observation and it considers behaviors in more basic components. The avoidance of supportive, nonclassroom behaviors also reduces the complexity of behavior observed. It avoids the possibility of erroneous inferences about high quality supportive behaviors necessarily resulting in effective teaching.

One cited advantage related to the objectivity of low-inference observation[11] is that if results of observation are used for certification or related processes for retention or exclusion from the profession, low-inference observation is not likely to be challenged in the courts. At least, the objectivity of it is not likely to be challenged.

Even with its objectivity, low-inference observation has some disadvantages. Opponents would challenge its limited scope by focusing on certain classroom behaviors of teacher and students. This disagreement is probably more a disagreement in philosophy about teaching, the generality-specificity issue in teaching, or perception of the nature of teaching and learning. Nevertheless, in view of the present lack of consensus among research results on effective teaching, it is important to retain the option of measuring what the profession (or segments thereof) perceives to be important.

At a more operational level, low-inference observation has some disadvantages associated with the scoring of results. Since there is no qualitative measure to the data as recorded, it is necessary to use frequencies which often come from different numbers of observations. Also frequency of occurrence may be greater for some items or cells than others. So there is usually a standardized procedure used so items contribute equal weights. Thus, the scores generated are relative scores. Although such scores are useful when considering individual teachers (diagnostic purposes, for example), they have definite limitations for group research when the data base is limited or when one group dominates the data base. Additional conceptualization and development of scoring procedures is definitely desireable for low-inference observation.

---

[11] Donald M. Medley, "The Role of Teacher Educators in Teacher Competency Assessment," COMPETENCY ASSESSMENT IN TEACHER EDUCATION: MAKING IT WORK, CONFERENCE RESOURCE BOOK, (Compiled by Donna Gollnick, Sharon McMullen and Karl Massanari) (Washington, D.C.: American Association of Colleges for Teacher Education, Nov. 1980), pp. 1-60.

The Coker as a Low-Inference Instrument. The low-inference observation instrument used in this research was the *Classroom Observations Keyed for Effectiveness Research* (COKER). The COKER is an objectively administered, low-inference system for recording teacher/student interactions and transactions in the classroom. The COKER is a relatively comprehensive system for recording student and teacher behavior as it takes place in the classroom and is appropriate for elementary or secondary level classes.

The COKER consists of a matrix of 389 cells in two sections: Section A is 2-dimensional and represents the interaction between teacher and students; Section B is 1-dimensional describing the teacher's and student behaviors. The COKER is the result of an extensive developmental process and is a composite instrument derived from five objective, systematic, direct observation instruments used in studies of teaching.[42] The items on the COKER have been generated and adapted from more than 1,300 category, sign items on these five different observation instruments. The intent of synthesizing items from the five instruments was to capture, to the extent possible, the breadth of the original instruments. Selection of items was based on correlations between scores on individual items and pupil gain and the collective judgment of experts. The instrument resulting from this process was called the *Georgia Assessment of Teacher Effectiveness* (GATE) which became with further, slight modifications, the COKER. The COKER is likely the most comprehensive and useful instrument available for low-inference observations of teacher behavior. A copy of the instrument is found in Appendix C.

---

[42] These instruments were the Observation Schedule and Record, Form 5 Verbal (OSCAR), Spaulding Teacher Activity Rating Schedule (STARS), Florida Climate and Control System (FLACCS), Teacher Practices Observation Record (TPOR) and Coping Analysis Schedule for Educational Settings (CASES).

Section A of the COKER is a matrix of numbered cells that designate specific teacher and student interactions. There are three major categories of teacher behaviors in this section: presenting, questioning, and responding. These categories are further subdivided into 37 subcategories. Ten of these subcategories involve student behaviors dealing with compliance to the teacher, asking and responding to questions, voluntary comments, and off-task comments and behaviors. Twenty-seven categories involve teacher behaviors, further subdivided into the three types of teacher behavior stated above. In teacher behavior where "presenting" takes place there are seven categories involving the initiation, preparation for, maintenance, and termination of classroom learning activities. In the "questioning" area of teacher behavior there are six categories which relate to the types and uses of teacher questions in instructing a lesson. In the "responding" aspect of teacher behavior there are fourteen categories designed to show ways in which a teacher responds to student work in terms of answers and questions.

Section B of the COKER instrument does not contain a matrix for recording behavior but does feature a record or occurrence of student and teacher behaviors. There are 33 categories of student behaviors. These contain both positive behaviors which serve to promote learning and negative behaviors which tend to disrupt learning. There are 86 categories of teacher behaviors. These are divided into four areas — methodology, affect and control, verbal and grouping. The methodology area alone contains 32 categories involving items dealing with ways of motivating students, types of student choice of class activities, ways of supervising students, ways of organizing (focusing) a lesson, sources of materials for instruction, aspects of student responsibility and evaluation, the encouraging or discouraging of student expression, the preventing or fostering of student perplexity, the correcting of student misinformation, teaching strategies, items relating to the cognitive level of the methodology, and finally, ways in which students spend their time. The affect and control aspect of the COKER has 21 categories reflecting the feelings of the teacher toward the students and the types of intervention used by the teacher to modify student behavior. Specifically, there are five items dealing with positive affect for reinforcement and control and 16 items relating to negative affect used for control. In the verbal area, there are 14 categories relating verbal behavior, positive and negative, for reinforcing and controlling students. Finally, in the grouping area there are 19 categories relating to ways of grouping students. Ten of these involve the group activities (work) under the control of the teacher while nine are associated with group activities in which students have major choice.

When using the COKER instrument at The University of Toledo, five two-person teams of observers are used and each team visits each student teacher twice per quarter. Each visit is 30-40 minutes in duration. Section A is recorded in the first five minutes of observational time and Section B is completed in the next five minutes. After an interval, the observation is repeated. Four sheets (front and back) or two Sections A's and two Section B's are completed during each visit. A total of 16 sheets (eight A's and eight B's) are completed by each team of two observers for each student teacher per quarter in The University of Toledo undergraduate student teaching program.

All data are collected by observers trained in the use of the COKER. With an observer team concept in use (five teams), there are ten COKER observers operating during each quarter. A reliability check is done on each team of two observers. This check is an attempt to insure agreement between the two observers. Practice sessions and simulations using video-tapes to record and compare behaviors are used. An outside evaluator is employed to establish reliability measures between the five teams themselves. Furthermore, practice sessions with active student teachers in local schools are used by the teams for observer agreement studies.

Quarterly meetings with observers are held to evaluate all phases of the operation — recording procedures, unclear items on the instrument, manual explanations, and suggestions for improvement of both the instrument and the manuals which accompany them.

The COKER being a low-inference observation instrument only requires the process of observing and recording teacher/student behaviors as they occur in the classroom; no more, no less. The incidence (or lack thereof) of individual behaviors, as checked on the observer forms, are usually combined according to some type of scoring algorithm to obtain scores on competencies. This scoring is objective, it can be done by a clerk or a computer, and no inferences are required in translating the observation data into scores once the algorithm has been defined. The COKER is scored using a standardized procedure across several items which comprise a "key." The key may represent a teacher competency, a more global teacher behavior, or a teaching strategy (in whole or in part). For the purposes of previous research and this research, scoring keys have been developed for The University of Toledo Teacher Education Program competencies. The scoring procedures for the COKER will be further discussed in a later section of this chapter. A "Users Manual" for the COKER has been published (1982) and is available from Joan G. and Homer Coker, P.O. Box 1017, Carrollton, Ga. 30117.

High-Inference Observation

High-inference observation in classrooms requires an observer to make substantial inferences as to what teacher behaviors are being demonstrated and the quality of such demonstration. As with low-inference observation, data can be obtained for competencies or groups of behaviors. Observers, usually working from a set of indicators and descriptors for the competencies (or whatever is being observed), make judgments on particular aspects of teacher performance. The actual procedure utilized is explained in the description of the use of the Teacher Performance Assessment Instruments (TPAI).

The concept of including a qualitative judgment about teacher behavior is intuitively appealing, and it has a practical advantage in the scoring procedure over low-inference observation. Scores can be generated on a scale of measurement that at least approaches an interval scale. Scores can be considered in this way in an absolute sense and comparisions can be made across competencies.

The major criticism of high-inference observation centers on the subjectivity which can enter in various ways. The subjectivity of the quality of what is being observed is the most obvious source. Opponents of high-inference observation often suggest that an individual's score is more a function of who does the observing than what actually occurred. This criticism may be countered, at least in part, by a consideration of professional expertise. Consider an example from the medical profession. Two medical students may both be competent to make incisions and to put in the sutures for a defined surgery. Yet the quality of the surgery performed may differ, and we generally accept the notion that a trained surgeon can distinguish the quality. It is similarly claimed in education that trained observers should be able to distinguish quality when observing the teaching-learning situation.

Another factor of subjectivity is introduced if an absolute "standard" is imposed for judging performance, such as a mean of 3.0 (using the 1-5 scale) across the behaviors for a competency. Certainly, this would be a subjective judgment; however, as data bases are increased, normative data could be used if desired.

High-inference observation instruments can include non-classroom indicators, although it is not a requirement that they be included. For the instrument used in this project, such indicators were available.

In summary, high-inference observation has its advantages and disadvantages. With the present state of the art of the measurement in this area, it would be unwise to summarily dismiss either high-inference or low-inference observation. Both were included in this research effort because each apparently had something to offer in improving competency measurement.

The TPAI as a High-Inference Instrument. The high-inference instrument used for this study was the *Teacher Performance Assessment Instruments* (TPAI). This instrument from the University of Georgia was developed through an extensive review and consensus process. It was created in the context of the competency-based teacher education program at the University of Georgia and the decision of the Georgia Department of Education to employ teacher competence considerations in teacher certification. Hundreds of teaching competency statements from a variety of sources were collected, sorted and synthesized. A survey instrument consisting of 52 competency statements was administered to a stratified sample of 4,668 Georgia educators. Only competencies consistently classified as generic and essential were retained resulting in 22 competency statements for inservice teachers and 17 for student teachers. The TPAI is now used in Georgia, providing evidence through observation for certification.

The TPAI contains five categories of competency statements called instruments. They are:

Teaching Plans and Materials (TPM)
Classroom Procedures (CP)
Interpersonal Skills (IS)
Professional Standards (PS)
Student Perceptions (SP)

The final category is not contained in the student teacher version of the TPAI. In fact, this category is not part of any observation but consists of a 30-item paper and pencil rating scale to which a teacher's pupils respond.

The four instruments involving observation include 51 behaviors. These instruments are individually discussed below.

The Teaching Plans and Materials (TPM) instrument consists of 15 indicators of a teachers competence in planning instruction and selecting procedures. It is divided into five sections as follows:

1. Plans instructions to achieve selected objectives.
   This section contains five items relating to the planning of objectives, teaching procedures, content, materials of assessment and instruction at varying levels.
2. Organizes instruction to take into account individual differences among learners.
   This section consists of three items relating to organizing instruction by considering capabilities, learning styles and rates of learning.
3. Obtains and uses information about the needs and progress of individual learners.
   This section has two items concerning the use of evaluation materials as well as communication with learners about their needs.
4. Refers learners with special problems to specialists.
   This section contains three items relating to using student records, identifying learners who need specialists and using information to assist learners with their problems.
5. Obtains and uses information about the effectiveness of instruction to revise it when necessary.
   This section uses two items relating to obtaining information about instruction and revising instruction as needed.

The Classroom Procedures (CP) instrument assesses the teacher's cognitive interaction with learners, skill in organizing and presenting instruction, and skill in classroom management. This instrument involves 20 behaviors and is divided into the following six areas:

1. Uses instructional techniques, methods and media related to the objectives.

This section contains three items concerning methods, aids and materials.

2. Communicates with learners.

   This section involves five items dealing with giving directions (explanations), using learner responses, providing feedback, and using written and oral expression.
3. Demonstrates a repertoire of teaching methods.

   This section consists of three items relating to the implementing of learning activities, demonstrating variety in methods, and demonstrating ability to use large and small groups as well as individual instruction.
4. Reinforces and encourages learner involvement in instruction.

   This section provides four items dealing with involving students in lessons.
5. Demonstrates an understanding of the school subject being taught.

   This section has two items relating to helping students to see the importance of subject matter and demonstrating knowledge of subject matter.
6. Organizes time, space, materials and equipment for instruction.

   This section uses three items concerning routine tasks, instructional time, and learning environment.

The third instrument involves Interpersonal Skills (IS) and involves ten behaviors. This instrument assesses the teacher's characteristics as they are related to climate in, and the teacher's ability to manage, a classroom. The instrument is divided into three areas as follows:

1. Demonstrates enthusiasm for teaching and learning and the subject being taught.

   This section contains three items relating to communicating enthusiasm, stimulating interest and conveying the impression of knowing what to do and how to do it.
2. Helps learners develop positive self-concepts.

   This section uses three items dealing with demonstrating warmth, sensitivity and empathy.
3. Manages classroom interactions.

This section has four items dealing with providing feedback on behavior, promoting interpersonal relationships, maintaining classroom behavior and managing disruptive behavior.

The fourth instrument involves Professional Standards (PS) and assesses the teacher's relationship with colleagues, the acceptance of professional responsibilities and efforts to improve professional skills. The instrument is divided into two areas as described below.

1. Meets professional responsibilities.
   This section contains four items dealing with working with colleagues, following procedures, demonstrating ethical behavior and performing extra-instructional duties.
2. Engages in professional self development.
   This section has two items dealing with participating in professional activities and sharing professional ideas and materials.

The above four instruments comprise the Teacher Performance Assessment Instruments developed by the Teacher Assessment Project, College of Education, of the University of Georgia. A copy of the instrument and a handbook for its interpretation can be obtained by writing Dr. William Capie, College of Education, The University of Georgia, Athens, Georgia 30602.

As has been explained, each instrument of the TPAI is divided into a specific number of sections properly called an "indicator." Each indicator has a scale of descriptors from 1 to 5. The greater the number of the descriptor the more complete is the descriptive statement about the achievement of the indicator making up the basic competency under consideration.

When using the TPAI each student teacher is visited twice per quarter by an observer who is usually the college student teacher supervisor. In addition to these two observations, the supervisor also conducts two interviews with the student teacher per quarter. All data are collected by these observers who are trained in the use of the TPAI. Such training was provided by the developers of the TPAI from the University of Georgia. The observers follow the directions for using the

TPAI as set forth in the training manual titled *A Handbook for Interpretation.*[13]

The Measurement of Competencies

However a competency is stated, it is operationalized by the measuring instruments used to obtain information about it. As has been indicated, both low-inference (COKER) and high-inference (TPAI) observation procedures have been used in the research/evaluation program at The University of Toledo. The decision to use the COKER and TPAI was not arrived at on an *ad hoc* basis. A thorough review of available instruments was conducted, and, in fact, some other instruments in a pilot project were tried out which proved unsatisfactory. It was decided early not to develop locally derived instruments because of the enormity of the task.

Mapping of the Behaviors Into the Competencies Measured by the Instruments

Adapting existing instruments introduces the question, "How well do the instruments measure the competencies?" This requires a mapping of the behaviors measured by the instruments into the competencies. Again, this is a task which requires synthesis, successive approximations, and operationalizing descriptive terms in the context of the instruments. Essentially, the mapping process results in identifying the behaviors which contribute to a competency. However, the process also contributes to the formulation of competencies. If, for example, two competencies consist of the same, or very nearly the same behaviors, then the two competencies are identical as far as measurement is concerned, regardless of what the competencies were called initially. Such a mapping round of competency analysis resulted in 85 of the original 96 University of Toledo competencies being synthesized into 49 competencies. (Eleven competencies were discarded as not being adequately covered by the instruments.) Of the 49 competencies established for The University of Toledo's CBTE Program, 45 were measured by the TPAI, 22 by the COKER, and 18 were common to both instruments. The 49 competencies were organized for convenience under the topics below and these appear in Appendix A.

---

[13] William Capie, S. Joy Anderson, Charles E. Johnson, Chad D. Ellett. TEACHER PERFORMANCE ASSESSMENT INSTRUMENTS: A HANDBOOK FOR INTERPRETATION. (Athens: Ga., Teacher Assessment Project The University of Georgia, 1980.)(Also obtained through Division of Staff Development, Georgia Department of Education, Atlanta, Ga.)

Planning, Teaching Materials/Equipment and Evaluation
Instructional Strategies, Techniques and/or Methods
Communication With Learners
Learner Reinforcement - Involvement
Professional Standards

The categorization of the competencies into the above headings were merely a device of convenience for organizing the competencies. The mapping of the behaviors into the competencies measured by the instruments resulted in the establishment of scoring keys, one key for each competency measured by each instrument. Thus, a scoring key consists of the behaviors (cluster of items) that represent a teacher competency. Scoring keys were established by faculty members engaged in the competency measurement process at The University of Toledo with the aid and supervision of the originators of the COKER and TPAI instruments.

The Scoring of COKER and TPAI Data

The data from the COKER and TPAI produce different types of scores on competencies. The TPAI produces a scaled score, on a scale of 1 to 5, and the COKER is based on tabulations of occurrence or lack of occurrence of behaviors.

The scoring of the TPAI is straightforward and typical of that used with any instrument involving a scaled scoring procedure. With the TPAI, an observer works from a set of indicators for each competency and makes a judgment as to how well the indicators are being performed. This judgment is based on the utilization of a set of five descriptors for each indicator. The observer simply picks the number of one of the descriptors, (1-5) and that number becomes the score for the indicator. The scores for the indicators of the competency are summed and a mean computed to adjust for the varying numbers of indicators across competencies.

The scoring system for the COKER is much more complicated since the original data do not contain a scale. The lack of scaling in such low-inference observation data introduces a complexity not found in scoring high-inference data. Only the incidence (or lack thereof) of behaviors are recorded. There are two scoring procedures for the COKER as described below.

One system for the COKER involves computing an average frequency for each item (cell) by classroom over however many observations are involved. Then the scores on each item are standardized to a distribution of z-scores. This establishes item weighting. The next step involves generating scores on competencies, that is, a key score. The standardized item scores are combined for a class or teacher, and the key scores are then standardized across all classes (teachers) to a distribution of 50 and a standard deviation of 10 (T-score), using the scores for all classes being scored. This transformation eliminates negative scores.

The above scoring process produces relative rather than absolute scores. For whatever number of teachers or classes are scored, the overall mean on the competency is always 50. Groups and individuals can be positioned in the overall distribution. Any change in the overall group requires a reanalysis of position. Also, this scoring system masks the nature of competencies with respect to frequent and infrequent occuring behaviors.

A second scoring system for the COKER is called the "proportionality score" which is not dependent on relative position for the computation of the score. Further, the proportionality score retains differences in competencies due to infrequently occurring behaviors. The proportionality model for scoring COKER data provides a value for each item or cell that maps into a competency, which is the proportion of times (p) the item was checked over how many times the teacher was observed. These proportions are then summed across the items of the competency. Since competency scores are based on a varying number of items, this sum is divided by the number of items in the competency, so that scores can be compared across competencies. Scores so generated can take on values from 0 to 1.0 inclusive. These are multiplied by 100 simply to provide a more familiar scale.

Items in the COKER are scored in one of two directions: the presence or absence of behavior. For those items scored on absence of behavior, 1-p rather than p is the value initially obtained. By doing this, the intent of the competency score is consistent across items.

In the following section advantages and disadvantages of the two scoring procedures for COKER data are discussed. Both approaches to scoring have been used with University of Toledo data. However, all COKER results of profile analysis and multivariate analysis reported on in this monograph are based on the proportionality scoring procedure.

The Nature of Compentency Scores

After the scoring procedure has been established, the question might be raised, "What do we have now that we have a score?" The TPAI, competency scores seem straightforward enough; they are scaled from the scores of the descriptors contributing to the competency. The characteristic of having a scale is certainly desirable for comparison and manipulation of scores. It makes the competency scores seem like scores on an achievement test, for example. But the scaling in the TPAI scores is obtained from the ratings (using the descriptors) given by the observer. So, the inherent validity of the scores rests upon the capability of the observer to make valid judgments.

There is no absolute basis for establishing "standards" of performance for the TPAI scores. With continued use there have been some empirical bases for expected performance, as data from "normative" groups have been obtained. However, there is no reason to believe that the expected performance for a defined group of student teachers (or teachers) will be the same across the 45 competencies measured by the TPAI. As profiles are interpreted later in the next chapter, it is important to keep this in mind.

The COKER scores are different that the TPAI scores; the COKER being inherently based on a dichotomy of occurrence-nonoccurrence of behavior, not on ratings. The score, regardless of whether the standardized T-score or the proportionality score is used, represents a quantification of occurrence or lack of occurrence of behaviors; it does not contain a qualitative meaning. If two student teachers have scores of 45 and 35 respectively on a competency it means that the former has (1) demonstrated more often the behaviors that are indicators of the competency, (2) demonstrated less often behaviors which indicate lack of the competency, or (3) a combination of (1) and (2).

The proportionality model has some advantages over the standardized T-score model for scoring. It has a conceptual zero point. The nature of individual competency scores is not masked as it is when scores are transformed to a standard score. For example, consistently low scores across teachers on a competency may indicate that the competency includes numerous items, scored on presence of behavior, which are seldom if ever checked. The proportionality model also provides some indication of absolute scores, rather than being limited to relative position interpretations. The standardized T-score has an advantage in that comparisons can be made across competencies since the scores are in standard form.

The COKER scoring keys for competencies vary considerably in the number of items per competency and also the combinations of numbers of behaviors that are scored on presence and absence of behavior. These characteristics affect the scores on the competencies. Most competencies have large enough numbers of items, and there were adequate observations taken, that the possible scores are not unduly restricted. The proportionality score does reveal certain characteristics of competencies that tend to be masked, if some type of standard score system of scoring is used.

A single observation with the COKER is of relatively short duration and there is a limit to the number of behaviors that can be observed in such a time frame. Thus, the magnitude of the score on a competency is to some extent a function of the numbers of items scored on presence or absence of behavior. In order to illustrate the nature of the competency scores using the proportionality score, some examples are discussed below. These are based on the data from the profiles of Chapter 4.

C-33, "Gives clear directions and explanations," is scored on 16 items, eight each on presence and absence of behaviors. The mean on this competency is around 60. C-34, "Motivates students to ask questions," is scored on ten items, eight of which are scored on presence of behavior. The mean on this competency is approximately 26. The competency C-34 has more items scored on presence of behavior. Undoubtedly, at least part of this lower score is due to the limited time for demonstrating the necessary behaviors. C-33, on the other hand, has its score increased by the failure of those behaviors to appear which are scored on absence of behavior.

Consider more extreme examples. C-22, "Structures the use of time to facilitate student learning," is scored on 12 items, all on absence of behavior. The reason for the extremely high mean of around 98 is that practically none of these behaviors are observed. On the other hand, C-24, "Provides learning experiences which enable students to transfer principles and generalizations to situations outside of school," has 13 items all scored on presence of behavior. The mean for C-24 was around 6, indicating that the behaviors for this competency seldom appear.

So what should be concluded from these competency scores? When the magnitude of the score is considered with the number and type of items for the competency, the competency is inherently described. There may be behaviors which are seldom demonstrated but, nonetheless, very important for teaching. We would not discard C-24, for example, just because it has a low mean. A low score might very well be expected for that competency. To some extent a competency such as C-22 with its consistently, very high scores near the maximum score is a "ghost competency." But it has value in that if for any group or individual the score drops substantially, say around 10-15 points, it is an indication of considerable undesirable behavior, behavior that shows a lack of this competency.

One point that should be noted about the COKER proportionality scores is that they should not be compared directly across competencies. For example, we would not conclude that the student teachers were performing poorly on C-34 just because its mean is lower than the mean of C-33. The scores are a function of the nature of the competencies, namely the behaviors that make up the competencies. If comparisons are to made across competencies, the standardized T-score is the better score, although it masks the nature of the competencies.

Validity and Reliability

The 49 competencies of The University of Toledo program were based on a content analysis of the program modules. Thus, in relation to the program, the competencies have content validity. The correspondence between the competencies of the program and the competencies identified as essential for enhancing learning in the statewide survey was discussed in an earlier chapter. This correspondence provided further evidence of the face validity of the competencies.

Shifting the focus of validity to the inventories as they measure competencies, again validity is essentially "face" or content validity; the extent to which the items, indicators and descriptors of the systems reflect the competencies being measured. Validity depends upon the correspondence between items and competencies, and an item has only its content to represent it.

Theoretically, construct validity could be considered, focusing on the constructs which underly teaching. Factor analysis would be an approach to establishing construct validity. A subsequent chapter deals with multivariate analyses including the results of factor analysis computed on the observation data. However, these factor analyses were done as exploratory and descriptive analyses. There were not *a priori*, "Theories of Teaching," so to speak, tested by factor analyses. Those analyses were not based on any hypotheses about the numbers of constructs or the grouping of competencies. So, although the empirical results of factor analyses provide information about possible constructs underlying the competencies, they should not be interpreted as the construct validation of any theory.

Some authors would argue that the validity of competencies should be established through their relationships to student achievement. Although this idea is intuitively appealing, its operationalization is very difficult. The results in this area are far from conclusive, and as Medley (1977) has indicated, may be situation specific with interactions operating. So, for the present at least, validity seems to be primarily content validity.

Reliability is the reliability of the competency score. Also, observer agreement often is referred to as a type of reliability, such as interobserver reliability. Observer agreement is important and is estabished through observer training. It is essentially a necessary but not sufficient condition for reliability of competency scores. A great deal of attention has been given to reliability of both the TPAI and the COKER. Brief comments are made below, and documents indicated are references which discuss reliability in greater detail.

Reliability of the TPAI

The developers of the TPAI have conducted numerous reliability studies, both with beginning teachers and student teachers. Capie, *et al* (1979)[44] provided a summary of results. Cronbach Alpha internal consistency coefficients were computed for the ratings on each of the three major instruments making up the TPAI. A number of raters, varying from 11 to 16, provided ratings in third grade science and eighth grade language arts. Teaching Plans and Materials, Classroom Procedures and Interpersonal Skills were the instruments rated. The Cronbach Alpha coefficients are presented in Table 1.

Table 1

Cronbach Alpha Coefficients for Three Instruments of the Student Teacher Version of the TPAI[45]

Instrument	3rd Grade Science	8th Grade Language Arts
Teaching Plans & Materials	.80	.75
Classroom Procedures	.75	.88
Interpersonal Skills	.65	.44

The coefficients for Interpersonal Skills were somewhat lower than those for the other two instruments. This may be accounted for by the fact that the variance within ratings for Interpersonal Skills was less than variance for Teaching Plans and Materials and Classroom Procedures.

Concerning interrater agreement on indicators of the TPAI, the developers of the TPAI used two categories, exact agreement, and near agreement. Coefficients of near agreement tended to be in the .70s and .80s with some in the low .90s. As expected, exact agreement coefficients were somewhat less. But this level of agreement was considered adequate. These same indicators were used for The University of Toledo competencies, and it is inferred that agreement would be similar. Correspondingly, competency scores are made up of the same items, and the reliabilities would likely be similar in magnitude as those found by the developers of the TPAI.

---

[44] William Capie, ET AL, SELECTED INVESTIGATIONS OF THE RELIABILITY OF THE TPAI. (Athens, Ga.: University of Georgia working Paper, Teacher Assessment Project, College of Education, The University of Georgia, 1979.)

[45] IBID. P. 38

Reliability of the COKER.

Reliability information, conceptual and empirical, is provided in the *Users Manual* (1982)[46] for the COKER, and the details will not be repeated here. The reliability of greatest concern is the internal consistency reliability estimated through the Cronbach Alpha coefficient. Essentially this is the internal consistency of the items which apply to a competency. Stability reliability may also be considered; however, maintaining stability across a series of observations may be of little concern, especially when observing student teachers. They may be teaching different subjects, especially in the elementary grades, or there may be direct attempts to change teaching performance.

Estimation of internal consistency of competency scores has some problems associated with it. For example, one competency, C-23, has only one item so the α - coefficient does not apply, and all reliability estimates for the score on this competency would be zero. The nature of the competency scores also affects reliability estimates. Large numbers of unchecked cells for a specific competency affect the estimates. A number of reliability runs were computed on the COKER scores, both for the proportionality score and the standardized T-score. Coefficients ranged from a high of around .80 to a low of zero for C-23. Competencies with small numbers of items tended to have the lowest reliability coefficients; those with the greatest number of items had the higher coefficients. Reliability estimation remains a problem when dealing with measurement through observation, not only empirically but also theoretically, as to which procedures are most applicable.

Summary

This chapter has focused on the methodology of classroom or teacher observation, and includes the initial point of identifying the competencies to be measured, a discussion of high and low inference observation, a description of the instruments used (COKER/TPAI), the acquisition of scores on competencies, what such scores represent, and information on validity and reliability. The entire process has several steps and is quite complex. Since teaching is a complex activity, it should be expected that measurement through observation is likewise complex. It is not likely that a less demanding, superficial process would generate useful results.

---

[46] Joan G. Coker and Homer Coker. CLASSROOM OBSERVATIONS KEYED FOR EFFECTIVENESS RESEARCH, USERS MANUAL. (Atlanta, Ga.: Georgia State University/Carroll County Teacher Corps Project, 1982.) pp. 13-19, 31-37.

The contrasts between high and low-inference observation were presented in this chapter. Certainly a major contrast is the qualitative meaning contained in a high-inference score, which is absent from the low-inference score. This characteristic does lend an apparent advantage of objectivity to low-inference observation. High-inference observation is not as restricted as low-inference observation. It can deal with supportive, but nonclassroom-demonstrated, behaviors. This is the primary reason why the TPAI measures 45 of the 49 University of Toledo competencies and the COKER measures 22 of 49. Be this as it may, each approach has enough to offer to merit its retention, at least for the present, in the evaluation of The University of Toledo student teachers and the CBTE program.

Thus far, the emphasis in this monograph has been upon reviewing the field and explaining the methodology, essentially setting the context for the evaluation of teacher performance. In subsequent chapters, analyses and results are discussed. Observation quickly generates a large data base, so analyses must focus on synthesizing data, especially when conducting program evaluation. The analyses described in the following chapters by no means include all possible types of analyses that can be conducted on the observation data. They are analyses that appeared to provide some of the most important information from the standpoint of student group and CBTE program analysis.

Chapter IV

PROFILE ANALYSIS AND RESULTS

Forty-nine competencies were identified for The University of Toledo (UT), CBTE program which are listed in the Appendix. The competencies were grouped according to five "topics," and this grouping was simply a way of organizing competencies. As a set of competencies, the UT competencies are quite similar to competencies that might be generated for any teacher education program, especially one that is competency-based. But the operational definitions of the competencies are determined by how the competencies are measured and how the data generated by the observation inventories are scored. These subjects were presented and discussed in the previous chapter. This chapter provides the next step in competency data analysis by examining competency scores through profiles.

Profile Analysis

Profile analysis is essentially the comparison of competency scores (represented by some statistic such as the mean or median). Profiles can be generated for groups or individuals, although the emphasis here is on those for groups. The scores on the TPAI can be compared across competencies if the assumption can be made that expected performance is about the same. This may be a tenuous assumption for student teachers. Nevertheless, score positions can be identified, and they do provide information for program evaluation. The COKER scores can be compared across groups, but for the reasons previously described comparisons across competencies must take into account the nature of the competencies.

TPAI Results

The data from 309 student teachers were separated into three teaching areas: elementary (134), secondary (122), and special education (53). Twenty-seven student teachers could not be uniquely identified according to teaching area. A profile of means for the three groups appears in Figure 2. The basic data for the profile appear in Table 2.

Profile of Mean Competency Scores for Elementary, Secondary and Special Education; UT Student Teachers. 1979-1981 - TPAI

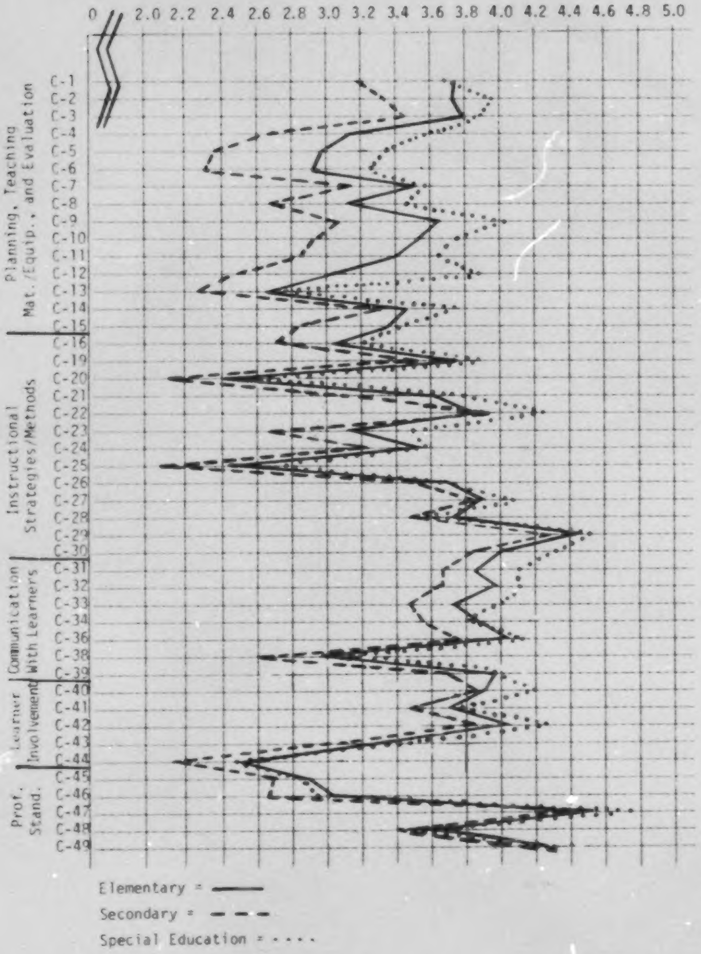


Figure 2

TABLE 2

Mean Competency Scores by Level Taught; UT Student Teachers; Fall 1979 - Spring 1981 - TPAI

Competency No.	Level Taught		
	Elementary	Secondary	Special Educ.
C-1	3.73	3.20	3.68
C-2	3.70	3.32	3.97
C-3	3.80	3.43	3.91
C-4	3.16	2.64	3.60
C-5	2.97	2.38	3.38
C-6	2.94	2.36	3.27
C-7	3.52	3.17	3.59
C-8	3.14	2.68	3.45
C-9	3.65	3.04	4.09
C-10	3.53	2.97	3.75
C-11	3.40	2.88	3.63
C-12	3.05	2.52	3.87
C-13	2.66	2.29	2.73
C-14	3.46	3.33	3.75
C-15	3.37	2.85	3.44
C-16	3.03	2.73	3.23
C-19	3.75	3.52	4.00
C-20	2.41	2.10	2.57
C-21	3.60	3.15	3.71
C-22	4.01	3.93	4.22
C-23	3.14	2.68	3.45
C-24	3.51	3.23	3.58
C-25	2.42	2.07	2.78
C-26	3.66	3.44	3.61
C-27	3.88	3.86	4.11
C-28	3.75	3.53	3.56
C-29	4.42	4.21	4.54
C-30	4.08	3.85	4.16
C-31	3.86	3.68	3.96
C-32	3.98	3.68	4.11
C-33	3.77	3.48	3.80
C-34	3.86	3.58	4.08
C-36	4.01	3.77	4.13
C-38	2.99	2.60	3.06
C-39	3.98	3.64	3.96
C-40	3.92	3.84	4.20
C-41	3.71	3.47	3.88
C-42	4.08	3.82	4.26
C-43	3.36	3.02	3.44
C-44	2.52	2.18	2.49
C-45	2.84	2.76	2.82
C-46	3.07	2.66	2.96
C-47	4.50	4.54	4.74
C-48	3.60	3.40	3.53
C-49	4.21	4.27	4.24

The patterns of means for the three groups are quite consistent across all competencies. This consistency is expected since much of the competency-based teacher education program provides common content, regardless of the teaching area. Certainly, program focus is fundamentally the same across teaching areas.

The greatest differences among the groups occur in the first 15 competencies which deal with planning, the knowledge of teaching materials/equipment, and evaluation procedures. In such competencies, the secondary group is consistently low and the special education group similarly high on the 15 competencies rated. Elementary student teachers are more like the special education group than secondary student teachers. Secondary teachers were particularly low on C-5, "Gathers, organizes and evaluates pertinent information about students for effective instruction," C-6, "Identifies and evaluates learning problems of students in the content area being taught," and C-13, "Engages in professional development by obtaining and analyzing evaluative information concerning the effectiveness of instruction." Apparently, secondary student teachers in their professional coursework had less instruction in the use of evaluation procedures and were actually not demonstrating as much activity as the other groups in this area in their student teaching. The differences between the three groups in the first 15 competencies suggests an instructional program review of the work taken and accomplished by students in all three groups.

All three groups of student teachers had low means on C-20, C-25, C-38, and C-44. C-20 and C-25 are both concered with aspects of individualizing instruction. The first competency deals with modifications of instructional activities to accommodate identified learner needs and the second involves providing assignments and learning opportunities appropriate to different ability levels of pupils. The other two low scoring competencies, C-38 and C-44, concern evaluation skills and the provision of feedback to learners on their performance. These findings tend to fit the other low scores found on evaluation procedures in the first competency grouping. It may be that competencies in these areas are more difficult to develop than other competencies. In any event they merit additional attention in The University of Toledo CBTE program, especially competencies dealing with evaluation.

There were a number of competencies on which all groups had high scores. These were C-19, C-22, C-27, C-28, C-29, C-32, C-36, C-39, C-42, C-47 and C-49. C-19 through C-29 deal with instructional strategies, techniques and methods, specifically:

- C-19, "Establishes transitions and sequences in instruction which are varied;"
- C-22, "Structures the use of time to facilitate student learning;"
- C-27, "Demonstrates self-direction and conveys the impression of knowing what to do and how to do it;"
- C-28, "Works effectively as a member of an instructional team;" and
- C-29, "Uses acceptable written and oral expression with learners."

C-32, C-36 and C-39 involve communciation with learners and vary from using a variety of functional verbal and non-verbal communication skills to the expression of a positive, personal attitude toward teaching. C-42 involves the use of positive reinforcement patterns with students with the final two competencies, C-47 and C-49, concerned with the general topic of high professional standards. It is interesting to note that where all groups had high scores, certain and stressed aspects of The University of Toledo, CBTE program were reflected. For example, the competencies involved with varied transitions and sequences in instruction, the structuring of time to facilitate student learning, working effectively as a member of an instructional team, using a variety of verbal and non-verbal communication skills, and the use of positive reinforcement patterns with students are all distinct components of the professional CBTE program. Also reflected in high scores were C-2, planning instruction at a variety of cognitive levels, C-3, the use of behavioral objectives, and C-40, keeping students on task. These scores were additional indications of CBTE program emphasis. The high scores may also be due in part to competencies more readily developed with student teachers.

There were 27 student teachers (over the six quarters) that could not be identified uniquely as elementary, secondary or special education. The data from these student teachers were included with those of the 309 discussed above, and a profile of means and medians from all 336 student teachers was developed. This profile appears in Figure 3, and the corresponding information along with the standard deviations is given in Table 3.

Profile of Mean and Median Competency Scores for UT Student Teachers; 1979-1981 - TPAI N = 336

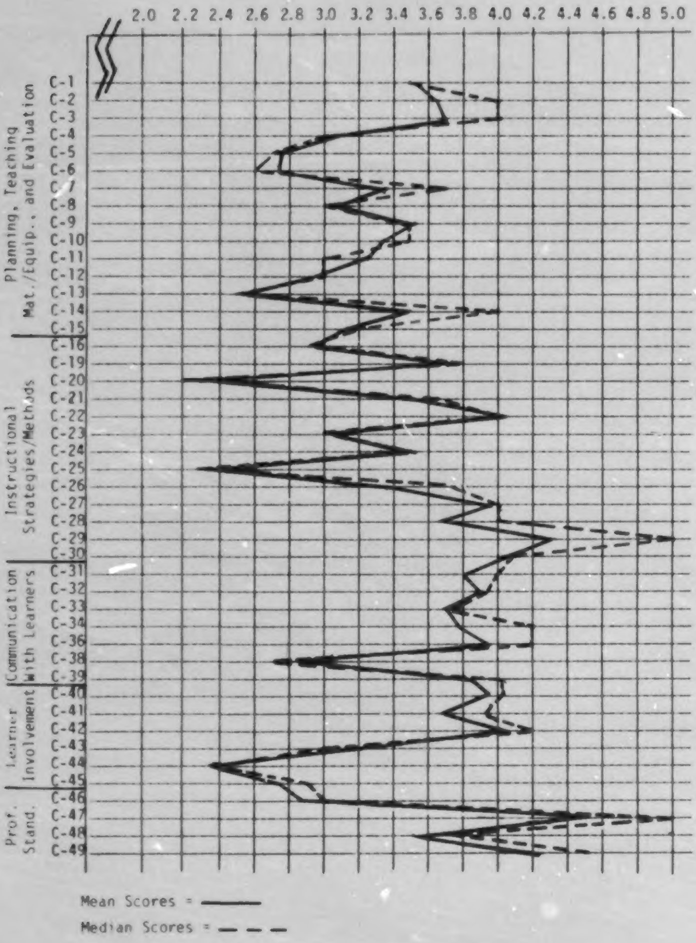


Figure 3

Table 3

Means, Medians and Standard Deviations for Competency Scores; UT Student Teachers, Fall 1979 - Spring 1981 - TPAI (N = 336)

Competency No.	Mean	Median	S.D.
C-1	3.53	3.50	.81
C-2	3.62	4.00	.91
C-3	3.68	4.00	1.00
C-4	3.03	3.00	1.02
C-5	2.78	2.71	.94
C-6	2.73	2.60	.93
C-7	3.38	3.67	1.07
C-8	3.03	3.00	.88
C-9	3.51	3.50	1.09
C-10	3.36	3.50	.95
C-11	3.24	3.00	1.01
C-12	2.96	3.00	1.01
C-13	2.50	2.50	.82
C-14	3.48	4.00	.92
C-15	3.16	3.20	.88
C-16	2.94	2.93	.64
C-19	3.71	3.80	.80
C-20	2.29	2.20	.73
C-21	3.45	3.63	.74
C-22	4.02	4.00	.82
C-23	3.03	3.00	.88
C-24	3.44	3.50	1.04
C-25	2.33	2.25	.76
C-26	3.58	3.63	.90
C-27	3.93	4.00	.83
C-28	3.63	4.00	.94
C-29	4.37	5.00	.81
C-30	4.02	4.17	.80
C-31	3.80	4.00	.89
C-32	3.89	3.96	.68
C-33	3.67	3.75	.77
C-34	3.78	4.00	.85
C-36	3.94	4.00	.79
C-38	2.84	2.75	.74
C-39	3.85	4.00	.90
C-40	3.94	4.06	.81
C-41	3.66	3.78	.69
C-42	4.02	4.20	.79
C-43	3.23	3.17	.64
C-44	2.36	2.38	.75
C-45	2.79	2.92	.57
C-46	2.87	3.00	.85
C-47	4.54	5.00	.73
C-48	3.51	3.67	.94
C-49	4.21	4.50	.88

For most competencies, the mean and median are quite close. However, there are some interesting patterns among these statistics within competencies. Consider, for example, C-29, "Uses acceptable written and oral expression." At least one-half of the student teachers measured had scores of 5.0, the maximum score. The standard deviation was modest in size so the scores are "bunched" near the top of the scale. C-47, "Acts as an appropriate model in terms of ethics, attitudes and values," had the same pattern. C-38, "Provides feedback to learners on their cognitive performance," had a low mean and a low median, with a small standard deviation. Scores were consistently low for the student teachers on this competency. C-9, "Uses criteria and effective procedures for determining pupil achievement of learning objectives," had the mean and median very close and moderate in size, but it also had a large standard deviation, indicating considerable variability among the scores on this competency. C-20, "Modifies instructional activities to accommodate identified learner needs," had consistently low scores.

COKER Results.

The COKER observations involved 273 student teachers. Because of complications in the identification process, only 132 were identified for the three teaching areas: elementary (64), secondary (56) and special education (12). There were only 12 special education teachers observed because of the nature of instruction in most special education classes and those observations lost in the identification process. In classes where students are severely handicapped, the teacher behaviors and student behaviors are not adequately reflected by the COKER. A profile of means appears in Figure 4 and the corresponding data in Table 4. The proportionality model of scoring was used.

As had been discussed earlier in this chapter, the differences across competencies are due in part to the behaviors and their occurrences (presence or absence) which make up the competencies. For example, C-36, which tends to have the lowest mean has 25 items scored on presence of behavior plus two items scored on absence of behavior. The low mean apparently indicates that this competency consists primarily of behaviors that are recorded infrequently or rarely occur. On the other hand, C-22 which has the highest mean, is scored only on 12 absence of behavior items. For this competency the behaviors rarely occur and the "absence of behavior" items are infrequently checked, but because of the scoring direction the scores are very high

Mean Competency Scores by Level Taught; UT Student Teachers; Winter 1980 - Spring 1981- COKER

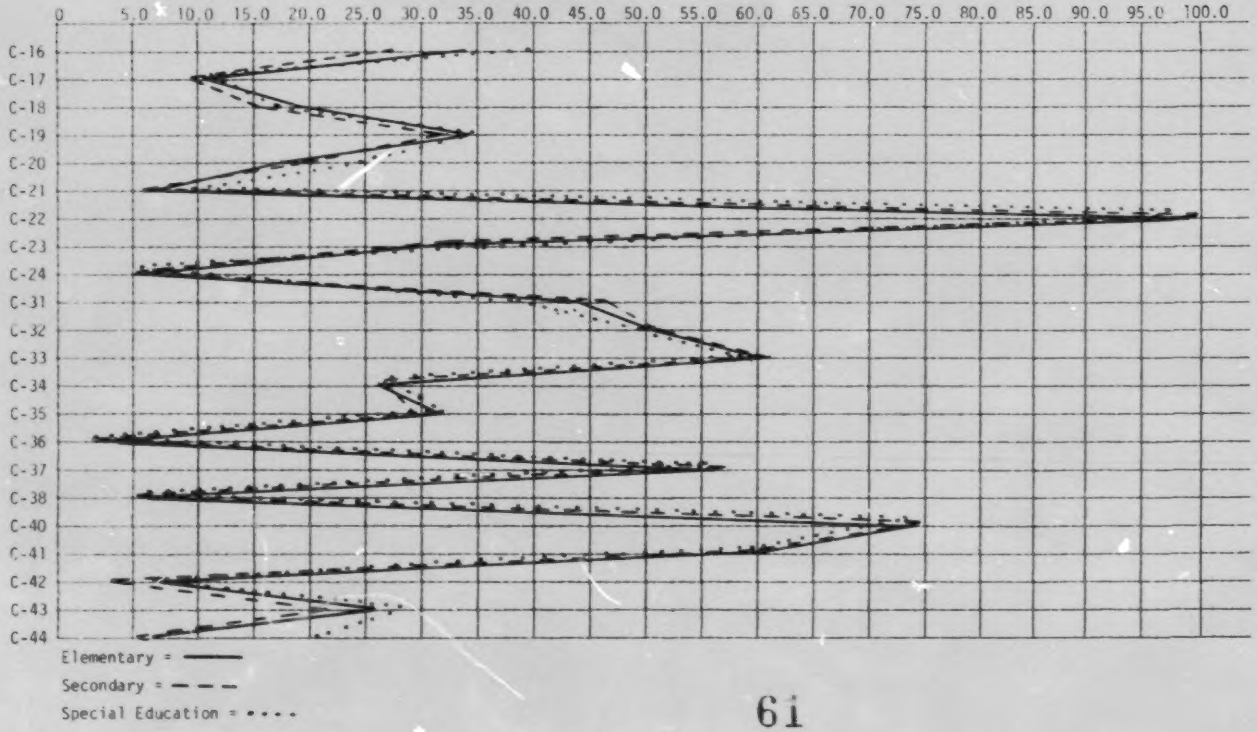


Figure 4

Table 4

Mean Competency Scores by Level Taught; UT Student Teachers; Winter 1980 - Spring 1981 - COKER

Competency No.	Level Taught Elementary	Secondary	Special Education
C-16	33.60	27.33	39.88
C-17	10.30	9.25	10.57
C-18	18.33	15.01	16.65
C-19	34.65	32.56	34.80
C-20	17.03	19.88	24.32
C-21	7.27	6.91	9.11
C-22	98.22	99.05	97.32
C-23	31.01	32.99	35.12
C-24	6.91	5.01	5.96
C-31	44.31	46.32	38.93
C-32	50.31	51.28	49.02
C-33	60.54	61.41	59.25
C-34	26.28	26.38	26.63
C-35	31.61	29.73	31.75
C-36	4.01	3.59	3.39
C-37	56.91	57.66	56.58
C-38	7.39	5.79	6.88
C-40	74.69	74.90	74.45
C-41	60.53	60.07	59.72
C-42	7.85	4.45	7.44
C-43	25.37	21.73	28.46
C-44	7.05	5.79	20.39

(i.e. the absence of behavior produces a high score). Competencies position themselves at various points on the scale of measurement, due in part to the pattern of items that map onto the competencies.

It must be remembered that COKER scores only have *quantitative, not qualitative,* meaning based on the occurrence or lack of occurrence of behaviors that make up a competency. Therefore, there can be no comparisons across competencies on the COKER as there can be when using the TPAI. No qualitative meaning as to the behaviors can be inferred with the COKER.

Scores across the three teaching groups on many of the competencies are quite similar. However, there are some differences among the groups, and the same group is not consistently low. Consider competencies C-16, C-20, C-43 and C-44. C-16, "Uses a variety of instructional strategies," shows secondary low and special education high with elementary about mid-way between the other two groups. C-43, "Assists students in discovering and correcting errors and inaccuracies," had the same pattern. On the other hand, C-20, "Modifies instructional activities to accommodate identified learner needs," had elementary low and secondary higher. C-44, "Develops student feedback, evaluation skills and student self-evaluation," had secondary and elementary means close together, but the special education mean was much higher. For these four competencies, the special education group consistently had the high mean.

The means for C-18, "Develops and demonstrates problem solving skills," were quite close, but what seems surprising is that on "problem solving" which should be stressed at the secondary level, secondary is 3.3 points below elementary. For C-31, "Provides group communication experiences for students," special education is low which is expected, but secondary not elementary is high. On C-42, "Uses positive reinforcement patterns with students," the secondary mean is lower than the means for the other two groups.

Using the data from all 273 student teachers, the profiles of means and medians were generated, and appear in Figure 5. Along with the standard deviations, the same information is contained in Table 5.

The means and medians are quite consistent for individual competencies. There are some exceptions such as C-23 and C-44 for which the differences are around five points. These competencies have scores

Mean and Median Competency Scores; UT Student Teachers;
Winter 1980 - Spring 1981- COKER

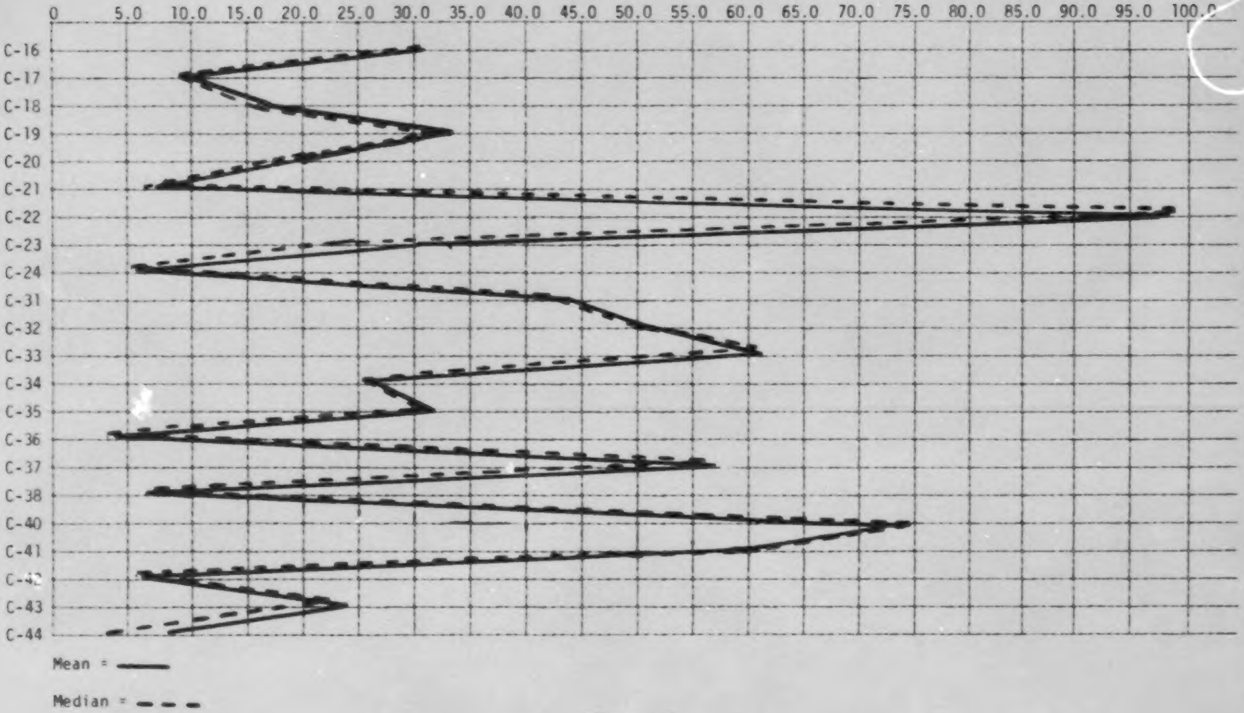


Figure 5

Table 5

Means, Medians and Standard Deviations for Competency Scores; UT Student Teachers, Winter 1980 - Spring 1981 - COKER (N = 273)

Competency No.	Mean	Median	S.D.
C-16	30.70	30.48	11.92
C-17	9.62	9.38	4.31
C-18	16.94	16.68	6.91
C-19	33.40	31.66	4.61
C-20	19.47	20.00	12.13
C-21	7.24	6.84	5.17
C-22	98.39	99.48	2.19
C-23	30.07	25.00	24.02
C-24	5.93	5.77	3.71
C-31	43.93	41.26	8.72
C-32	50.88	50.83	3.16
C-33	60.71	60.56	3.52
C-34	26.16	25.63	4.31
C-35	31.29	31.25	8.19
C-36	4.09	3.52	2.92
C-37	56.96	56.26	1.78
C-38	6.52	6.64	3.05
C-40	74.61	74.61	1.69
C-41	60.09	60.00	1.46
C-42	6.39	6.14	3.09
C-43	23.39	23.23	5.68
C-44	8.16	3.15	10.55

grouped toward the low ends of the distributions, with a few, very high scores. The standard deviations also tend to be some of the largest among the competencies. In fact, C-23 has the largest standard deviation, almost twice that of C-20 which has the second largest standard deviation.

At the present time there is no basis for empirically establishing what the means should be on each competency. *A priori* judgments can be made, but for each competency the nature of the competency must be considered. As the data base is increased and data from additional groups are obtained, it will be possible to make realistic estimates of expected performance. However, as an arbitrary point for the present study, any mean 20 or greater is assumed to reflect adequate demonstration of the competency. In this respect, it is interesting to note that competencies C-38 and C-44 which deal with different aspects of evaluation in the classroom have low means. This type of behavior is apparently not being exhibited in the classroom to the extent it should occur, which agrees with the results obtained from the TPAI. Overall, the scores indicate that most of the 22 competencies measured by the COKER and considered important to the CBTE program are being demonstrated in student teaching.

Summary

Being a competency-based program, the competencies of the UT program are highly behavioral in nature. This characteristic is partly the reason for having a relatively large number of competencies (49) as teacher education programs go. Generally, the competencies require the demonstration of some type of skills; this is true even with those competencies measured by the TPAI which focus on supportive activities such as planning. In that sense, evaluation is performance-based; it is not limited to knowledge about subject matter or teaching, although such knowledge is necessary. The analogy might be drawn with medicine in that a medical student may have a complete knowledge about the terminology, process, etc., associated with surgery, but the mark of a competent surgeon is being able to perform successful surgery. The mark of a competent teacher is being able successfully to conduct instruction.

Competencies are operationally defined by their measurement; the behaviors, indicators and descriptors which map onto the competencies. As will be discussed in later chapters, it appears that competency measurement is quite instrument dependent. That is, a score on a given competency depends to some extent on the instrument used to measure the competency. To a large extent, this may be due to the difference between high-inference and low-inference observation. Whatever the reason, the difference shows up in the nature of competency scores. The TPAI scores are scaled and have qualitative meaning. The COKER scores are not scaled; they represent the occurrence or lack of occurrence of behaviors. They have quantitative meaning but no qualitative meaning as to the behaviors being observed.

The COKER and TPAI provide viable approaches to measuring the underlying competencies intended from The University of Toledo teacher education program. Although measurement through observation is demanding, it is certainly feasible and useful results are obtained. The two instruments combined appear to provide adequate measurement of the competencies, and confidence in the instruments has been developed. Although the instruments have gone through an extensive developmental process, their large-scale and long-term use in the UT program has required additional developmental work. One apparent aspect of this development has been the mapping of behaviors and descriptors onto competencies. Another aspect has been developing an alternate approach to scoring the COKER data. The developmental work was essential for the application of observation to the evaluation effort.

Profile analysis provided a variety of useful information, not only about the relative performance of groups of student teachers, but also information about the nature of competencies. The positioning of the COKER scores on the scale of measurement is reflective of the types of behaviors indicative of the competencies. The competencies as measured certainly provide a "mixed bag." There are some competencies with very high or very low scores which seem to appear as ghost competencies, but that does not mean that they should be deleted. They may well represent important skills in teaching.

The TPAI provides scaled scores which can be compared to a "standard." Such standards would be subjectively established; however, they would likely be based on the results of some normative group. For use with student teachers, the most relevant normative group probably would be a group of inservice teachers that by some criteria have been assessed to be successful. Of cource, the ultimate criterion of success would be the relationship (hopefully, substantial and positive) between scores on competencies and student achievement (students of the teacher being measured). However, the "state of the art" of research in this area is still a substantial distance from establishing this type of criterion. It may never be a feasible criterion to use with student teachers because their duration in actual teaching is quite brief with many external factors operating beyond their control. Possibly the most realistic approach with student teachers will be to compare them with normative groups of inservice teachers who have met the criteria.

A score of 3.0 is the midpoint of the TPAI. There have been suggestions of using 3.5 as the score for minimal competence, although such standard setting is strictly subjective. An inspection of the means for the 45 competencies measured by the TPAI shows several below 3.0. These lower means tend to be concentrated in areas of evaluation, individualization and variation of instruction. Possibly competencies in these areas are more difficult to develop than those in other areas. It may be that competencies in these areas are represented by more concrete behaviors, and hence their assessment is more distinctive. Or in some contexts, it may not be necessarily desirable to obtain a high score. For example, the secondary group had the low mean on C-16, "Uses a variety of instructional strategies." However, in a specific secondary classroom on a given day, it may not be effective to vary instruction. The most effective teaching may occur by persisting with a single, instructional strategy.

The COKER data tended to reinforce the scores on the TPAI. For example, the secondary mean on C-16 is the lowest of the three groups. Competencies reflecting evaluation and individualization tended to have lower scores on the scale of measurement.

Generally, the data indicated that UT student teachers demonstrated the competencies intended from the program. However, even taking into account the nature of competencies, there was considerable variability in the scores (means and medians) across competencies. This characteristic is good, because if all scores were uniform, it would be difficult to counter the argument that this is an artifact of the measurement. Certain program components such as evaluation and individualization merit attention. It may be that the program is deficient in these components. Or if not deficient, improvement in these program components may take precedent over those in other components related to competencies with higher overall scores.

68

Chapter 5

MULTIVARIATE ANALYSIS OF OBSERVATION DATA

Any study with a large number of variables, such as the scores on the 49 University of Toledo competencies, would seem to lend itself to multivariate analysis. (Multivariate analysis here is used in the sense of multiple dependent variables, not for example, as multiple independent variables in a factorial analysis.) At least, there should be some data reduction through multivariate analysis. The analyses done on The University of Toledo observation data were exploratory analyses. They were not done for the purpose of theory testing or for checking some *a priori* pattern of difference between groups. Factor analysis and multivariate analysis of variance (MANOVA) were used.

Factor Analysis

The TPAI and COKER data were analyzed by separate factor analyses. A PROMAX, oblique rotation was used, since some correlation between factors was expected. The "Rule of 1.0" was used for factor retention. That is, any factor with an eigenvalue greater than 1.0 has greater variance than the original measures, in this case competencies. The subjective criterion of a factor loading of absolute value of .40 or greater was used for determining the competencies that loaded on the factors.

The number of competencies that underly a CBTE program has always caused a minor dilemma. To limit the number of competencies to a few tends to make them too global and fails to reflect the specificity of the program. To have a large number of competencies makes the list unmanageable, and essentially reduces competencies to behavioral objectives or something similar. Also, there will tend to be substantial correlations between combinations of competencies unless that number is reduced to one. So, factor analysis is an empirical approach for synthesizing the relationships among competencies.

Factor analysis is also helpful in identifying the number and nature of the competency "dimensions" that underly a CBTE program. Of course, the extent to which it does this depends upon the inherent validity of the competencies for the CBTE program. Presently the research base in the area of competency identification and relevance is not adequate to generate a complete "competency theory," so to speak.

That is, a complete construct of necessary competencies has not been developed (either conceptually or empirically) so that the theory can be tested through factor analysis. If this were the case, factor analysis could be used to check the construct validity of a list of competencies. But factor analyses can provide empirical evidence about constructs within the context of the competencies being analyzed.

TPAI Results

The factor analysis for the scores on the 45 competencies measured by the TPAI generated six factors. The first factor was a very dominant factor, with 19 competencies loading on it, having an eigenvalue of 25.5 and accounting for about 57 percent of the variance. The second factor had an eigenvalue of 3.7, thus accounting for only 8.2 percent of the variance. Eight competencies loaded on this factor, two of which (C-16 and C-43) had also loaded on the first factor. Of course, subsequent factors extracted accounted for decreasing amounts of variance. The remaining four factors accounted for an additional 16 percent of the variance. The results of the factor analysis appear in Table 6. The factor names are subjective descriptions based on the competencies loading on the factors.

The dominant first factor has the characteristic of being somewhat of an omnibus factor. This is an undesirable characteristic for distinguishing constructs. It indicates that as the competencies are measured by the TPAI, they represent some global, general skill of maintaining a positive teacher-student environment. There was also considerable correlation between pairs of factors, most of these correlation coefficients being around .35 and the largest coefficient being .56.

As the factor analysis was computed for the TPAI, all factor loadings for a given factor had the same sign. The first and third factors had all factor loadings positive; the remaining four factors had all loadings negative. The highly positive and negative loadings define the respective ends of the underlying factor. Factor names are given a positive connotation. For those factors with negative loading we can multiply the loadings by -1 to obtain a pattern of positive loadings to coincide with the positive or favorable end of the factor as stated.

The stong first factor may imply that observers tend not to differentiate clearly between the competencies on the TPAI. Although persons who are strong on one competency may also be strong on

70

Table 6

PROMAX Factor Pattern - TPAI

	Loading	Factor Name
I		**Positive Teacher-Student Environment** λ = **25.5**
C-16	.446	Use a variety of instructional strategies.
C-19	.856	Establishes transitions and sequences in instruction which are varied.
C-21	.672	Demonstrates ability to work with individuals, small groups, large groups.
C-22	.799	Structures use of time to facilitate student learning.
C-24	.533	Provides learning experiences which enable students to transfer principles and generalizations to situations outside of school.
C-26	.597	Demonstrates knowledge in the subject area.
C-27	.856	Demonstrates self-direction and conveys the impression of knowing what to do and how to do it.
C-30	.796	Adjusts components of the physical/learning environment over which the teacher has control to facilitate learning.
C-31	.844	Provides group communication (cooperation, interaction, learning from others) experiences for students.
C-32	.804	Uses a variety of functional verbal and non-verbal communication skills with students.
C-33	.779	Gives clear directions and explanations.
C-34	.892	Motivates students to ask questions.
C-36	.801	Accepts varied student viewpoints and/or ask students to extend or elaborate answers or ideas.
C-39	.778	Expresses a positive personal attitude toward the teaching profession.
C-40	.944	Maintains an environment in which students are actively involved, working on-task.
C-41	.885	Implements an effective classroom management system for positive student behavior (discipline).

C-42	.898	Uses positive reinforcement patterns with students.
C-43	.522	Assists students in discovering and correcting errors and inaccuracies.
C-45	.567	Accepts responsibility, is dependable.
II	**Student Self-Evaluation and Variety of Activities**	**λ = 3.7**
C-8	—.455	Knows how to select (or construct), organize and use appropriate instructional materials and equipment to facilitate learning activities.
C-13	—.553	Engages in professional development by obtaining and analyzing evaluative information concerning the effectiveness of instruction.
C-16	—.591	Uses a variety of instructional strategies.
C-23	—.445	Uses a variety of resources and materials.
C-25	—.817	Provides assignments/learning opportunities interesting and appropriate to different ability levels of pupils.
C-38	—.686	Provides feedback to learners on their cognitive performance.
C-43	—.533	Assists students in discovering and correcting errors and inaccuracies.
C-44	—.986	Develops student feedback, evaluation skills and student self-evaluation.
III	**Professional Relations**	**λ = 2.6**
C-7	.723	Keeps informed of current professional/subject area literature and curricular learning materials/resources available.
C-13	.625	Engages in professional development by obtaining and analyzing evaluative information concerning the effectiveness of instruction.
C-28	.926	Works effectively as a member of an instructional team.
C-46	.797	Evidences cooperation with others (teachers, administrators, support staff, parents, etc.) in planning and teaching.
C-48	.931	Attends teacher and other professional meetings.
IV	**Evaluates Student Learning Experiences**	**λ = 2.0**
C-4	—.673	Has realistic expectations for the learning process and student readiness for learning.

C-5	—.904	Gathers, organizes, and evaluates pertinent information about students for effective instruction.
C-6	—.904	Identifies and evaluates learning problems of students in content area being taught.
C-9	—.659	Uses criteria and effective procedures for determining pupil achievement of learning objectives.
C-10	—.563	Selects/develops appropriate assessment techniques.
C-11	—.550	Collects, quantifies, and interprets data from appropriate assessment instruments.
C-12	—.816	Maintains evaluation records.
C-15	—.641	Relates to accountability issues concerning responsibilities to students, parents, and the instructional process.
V		**Plans Instruction, Assesses Effects** $\lambda = 1.4$
C-1	—.658	Plans units of instruction.
C-2	—.569	Plans instruction at a variety of cognitive levels.
C-3	—.622	Can state pupil outcomes and/or student course objectives in behavioral terms (behavioral objectives).
C-9	—.501	Uses criteria and effective procedures for determining pupil achievement of learning objectives.
C-10	—.567	Selects/develops appropriate assessment techniques and instruments for instructional activities.
C-11	—.493	Collects, quantifies, and interprets data from appropriate assessment instruments.
C-29	—.473	Uses acceptable written and oral expression with learners.
VI		**Professional Conduct** $\lambda = 1.2$
C-45	—.399	Accepts responsibility, is dependable.
C-47	—.863	Acts as an appropriate model in terms of ethics, attitudes and values.
C-49	—.852	Understands and follows school law, policies, and procedures and their effects on teachers, and teaching, including professional conduct standards.

another, there is more correlation here than expected. Thus, there may be some "halo effect" or other effect operating. In any case, the factor analysis results shed some light on the empirical structure of the observation of teacher competencies using a high-inference instrument.

COKER Results

The factor analysis of the scores on the 22 competencies measured by the COKER generated six factors. The first factor, which had seven competencies loaded on it, accounted for about 31 percent of the variance. The second factor accounted for 13 percent of the variance, and variance percentages gradually decreased down to the sixth factor which accounted for about 5 percent. In total, the six factors accounted for 72% of the variance. The factor loadings along with factor names appear in Table 7.

The COKER factors had a more even distribution of accounting for variance than those of the TPAI. The COKER factor analysis did not load primarily on one general factor. The solution also showed that the six factors were quite independent; the average interfactor correlation was near .10.

The competencies measured with the COKER are all demonstrated in the classroom situation. The corresponding factors tend to reflect a greater number and more specific classroom activities than those of the TPAI. Also fewer competencies load on individual factors, and for the COKER there is little redundancy of competencies loading on more than one factor. In the context of the competencies measured, the latter characteristic is helpful in distinguishing the factors.

Comparisons of TPAI and COKER Factor Patterns

An inspection of the factor patterns for the TPAI and COKER reveal numerous differences. Although the TPAI measures over twice as many competencies as the COKER, the analyses generated six factors each. Correspondingly, fewer competencies load on the COKER factors.

The TPAI factors appear to be more general factors than those of the COKER. To some extent this would be expected because the COKER measures only those competencies demonstrated in the

Table 7

PROMAX Factor Pattern - COKER

	Loading	Factor Name	
I		**Problem Solving and Questioning**	**λ = 6.77**
C-17	.682	Uses convergent and divergent inquiry strategies.	
C-18	.841	Develops and demonstrates problem solving skills.	
C-24	.738	Provides learning experiences for transfer to situations outside school.	
C-33	—.467	Given clear directions and explanations.	
C-34	.716	Motivates students to ask questions.	
C-35	.669	Uses questions that cause students to think critically.	
C-36	.913	Accepts varied student viewpoints.	
II		**Matches Activities With Learners**	**λ = 2.84**
C-20	.721	Modifies activities to accomodate learner needs.	
C-21	.767	Able to work with individuals, small and large groups.	
C-23	.489	Uses a variety of resources and materials.	
C-40	.656	Maintains environment in which students are on-task.	
C-44	.695	Develops student feedback skills, student self-evaluation.	
III		**Flexibility in Instruction**	**λ = 2.24**
C-16	.637	Uses a variety of instructional strategies.	
C-19	.825	Establishes varied sequences and transitions in instruction.	
C-22	—.846	Structures use of time to facilitate student learning.	
C-33	.535	Gives clear directions and explanations.	
IV		**Communication; Organized Classroom**	**λ = 1.75**
C-32	.776	Uses variety of functional verbal and non-verbal communication.	
C-40	.544	Maintains environment in which students are on-task.	
C-41	.717	Implements an effective classroom management system.	

V		**Discipline Feedback, Reinforcement**	$\lambda = 1.10$
C-38	.690	Provides feedback to learners on cognitive performances.	
C-41	.722	Implements effective classroom management system.	
C-42	.556	Uses positive reinforcement patterns with students.	
VI		**Listens, Communicates With Learners**	$\lambda = 1.01$
C-31	.743	Provides group communication.	
C-37	.861	Demonstrates proper listening skills.	

classroom. However, the TPAI also measures within-classroom competencies, so a mixture of factors might be expected; those which reflect the more general aspects of teaching and those which focus on more specific classroom skills. The fourth, TPAI factor is quite specific (evaluates student learning experiences); however, this factor has a modest eigenvalue of 2.0.

There were 18 competencies measured by both observation inventories. As far as factor loadings are concerned, those competencies are spread through all six COKER factors, but they are contained entirely in the first two factors of the TPAI, being heavily concentrated in the first factor. In fact, there is at least one competency from each of the six COKER factors that loads on Factor I of the TPAI. Of course, part of this result is due to the dominant first factor of the TPAI.

The lack of grouping of competencies across the two factor patterns, indicates that the factor patterns have little in common. At first, this may seem like a surprising result. However, another empirical result discussed in the next chapter supports this conclusion. The correlations between the TPAI and COKER scores on the same competencies (those 18 measured by both inventories) were generally of zero magnitude. Knowing this, there would be little reason to expect the competencies to group across the two factor patterns. So, we have the commonly measured competencies centered in Factor I of the TPAI but scattered across the six COKER factors.

The results of the factor analyses emphasize the importance of the differences in competencies as operationally defined by the way they are measured. The factor patterns are different, not only in the loadings, but also in the type of factors that appear. This implies that if constructs or theories of teaching are going to be developed empirically, the measurement involved will need to be considered carefully. It appears that the constructs will be affected by whether low-inference or high-inference observation data are used. Hopefully, there would not be as many "theories" as there are inventories used to measure competencies. This point has probably not received adequate attention in past work with observation inventories. There may be more instrument dependence of measurement in this area than has been anticipated in the past.

The factor analysis of the COKER data provided a more distinctive factor pattern. The determination of the number of factors to be retained is an arbitrary decision, although the rule of 1.0 seems to be acceptable. If the "Scree-test" had been used as a criterion, fewer factors would have been retained in the factor patterns. In the Scree-test, factors are no longer retained when the graph of their eigenvalues levels off. Again, the Scree-test requires an arbitrary decision. However, with the present state of theory development and measurement in this area, it is probably advisable to retain as much information as possible, however tenuous it may be in exploratory analyses.

Multivariate Analysis of Variance

In the factor analyses, the data banks were analyzed without the separation of groups of student teachers. That is, the factors that were generated were based on the data across whatever groups of student teachers were included. With MANOVA, the approach was different in that different groups of students were identified and the focus of attention was on how the groups differed on the artificial variables generated. In MANOVA these artificial variables are called canonical variates, analogous to factors in factor analysis.

MANOVA is basically an inferential statistics procedure and can be viewed as an extension of univariate analysis of variance; the extension being that MANOVA includes two or more dependent variables in a single analysis. However, MANOVA also generates information, especially as related to the canonical variates, that can be used in a descriptive sense. The canonical variates "describe" the constructs that underlie the dependent variables.

In analysis of variance we consider how different groups account for variance in the dependent variable by analyzing the differences between the means of the groups. This concept is consistent with what is done in MANOVA, in that the scores of groups are analyzed for differences. But when two or more dependent variables are included simultaneously, differences between means are no longer considered but rather how the groups differ on the canonical variates.

The canonical variates, if two or more appear, are orthogonal. The first canonical variate generated in a MANOVA accounts for maximum variance and produces maximum separation between the groups. Subsequent canonical variates account, in order, for smaller differences between the groups. This process continues until there is no longer any statistically significant difference between the groups, or the degrees of freedom are exhausted, whichever occurs first. The number of degrees of freedom equals the number of dependent variables or k-l, where k equals the number of groups, whichever is smallest. In most analyses k-l is smallest.

The TPAI and COKER data were analyzed separately. In fact, different classification systems for determining the groups were used for the two data banks.

TPAI Results

Rather than include all 45 competencies measured by the TPAI, fifteen competencies were selected for the MANOVA. These 15 competencies were those that had the highest loadings on the factors of the factor analysis discussed earlier. (Generally, these were competencies with factor loadings of .8 or greater.) It was decided to use the reduced number of competencies because of the redundancy in the factor analysis among the competencies as measured by the TPAI.

The area of preparation, elementary, secondary, or special education, served as the independent variable. The numbers for these three levels were 57, 56 and 41, respectively. The data for the 57 elementary student teachers were randomly selected from the data bank for 1980-81. This was done in an attempt to somewhat equalize the numbers in the groups.

Since there were two degrees of freedom, at most two, statistically significant, canonical variates could appear, and two were found. Canonical variates are described in terms of the real variables with which they are correlated. So, as in factor analysis, competencies loaded on canonical variates. It was decided to retain any correlations of absolute value, .30 or greater. Those correlations or loadings are given in Table 8. The mean scores in standardized form on the canonical variates for the groups of the independent variable were as follows:

	V1	**V2**
Elementary	.357	—.412
Secondary	.532	.348
Special Education	—1.085	.098

The first canonical variate loaded on nine competencies, all having negative loadings. Since the special education group had the high, negative standardized score on the canonical variate, it meant that this

79

Table 8

Loadings on Canonical Variates[1] of Selected Competencies Measured by the TPAI; Area of Preparation, Elementary, Secondary and Special Education

Competency Number	Canonical Variate	
	1(.67)	2(.33)
C-5	—.625	
C-6	—.555	
C-12	—.798	
C-19	—.400	
C-25	—.483	—.371
C-34	—.365	
C-40	—.334	
C-41	—.375	
C-42	—.311	
C-44		—.319

[1] Canonical correlations given in parentheses.

group scored high on the loading competencies, a result supported by the profile analysis discussed earlier.

The first canonical variate could be called, "Evaluation with supportive classroom environment." The evaluation competencies were those with the greater loadings. The second canonical variate had only two competencies with loadings of absolute value of .30 or greater. The two competencies, C-25 (Provides assignments/learning opportunities interesting and appropriate to different ability levels of pupils) and C-44 (Develops student feedback, evaluation skills and student self-evaluation), are somewhat different. Again, both loadings were negative; therefore, the elementary group scored highest and the secondary group lowest on this canonical variate. However, the second canonical variate is relatively "weak," with the first canonical variate dominating the differences between the groups.

COKER Results

The 22 competencies measured by the COKER were included as the dependent variables for the MANOVA, and the groups were determined by area of preparation for teaching. There were three groups; primary, intermediate and secondary, and the coding for these groups was done at the time of observation on the COKER form. Two hundred, sixty-five student teachers were uniquely identified as preparing for one of the three areas. Special education student teachers were included with the appropriate level if it could be designated. This was not always possible, hence the slight slippage in the total number included. Secondary includes both junior high and senior high. There were 135 primary, 64 intermediate and 66 secondary.

Of most interest were the canonical variates generated by the MANOVA. Since there were only three groups, two statistically, significant canonical variates at most were possible, and two did appear. These canonical variates were orthogonal and, of course, they were extracted in the order in which they account for the most variance between the groups.

Surprisingly, relatively few competencies loaded on the canonical variates with correlations of .30 or greater. There were two such loadings on the first canonical variate and three on the second, and these loadings are given in Table 9.

Table 9

Loadings on Canonical Variates of the Competencies Measured by the COKER; Area of Preparation, Primary, Intermediate and Secondary

Competency Number	Canonical Variate 1(.68)	2(.37)
C-32		.600
C-36		.346
C-37		—.337
C-42	.796	
C-43	.365	

The first canonical variate loads heavily on C-42 and to a lesser extent on C-43. So, this factor could be called, "Monitoring student understanding with positive reinforcement." The second canonical variate might be called the "Classroom communication with students" factor. It involves the communication skills, and it also includes acceptance of student viewpoints.

As with the TPAI data, the mean scores in standardized form on the canonical variates were determined for the three groups of student teachers. These mean scores were as follows:

	VI	V2
Primary	.634	.113
Intermediate	—.356	—.632
Secondary	—.952	.383

The first canonical variate had positive loadings, therefore the primary group scored highest and secondary group lowest. The intermediate group mean was closer to that of secondary than that of primary. On the second canonical variate, secondary and primary were highest in that order for the positive loadings which were on C-32 and C-36. But C-37 (Demonstrates proper listening skills) had a negative loading which goes with the negative mean of the intermediate group. This is consistent with the result that the intermediate group had the highest mean on this competency.

There was no overlap of competencies with loadings of .30 or greater across the two canonical variates. In this sense the two canonical variates were quite distinctive. The magnitudes of the loadings were quite variable being slightly less than those of the first.

Comparison of Factor Analysis and MANOVA

Factor analysis and MANOVA are both multivariate analysis procedures, but it is important to emphasize that they do not provide the same results. Both procedures involve data reduction or synthesis, and both generate artificial variables called factors and canonical variates, and these artificial variables must be interpreted in the context of their relationships to the real variables.

Factor analysis synthesized the competencies into the constructs underlying the competencies across all groups of student teachers. The factors are probably most informative in a descriptive way. They "describe" the nature and skills of teaching reflected by the specific competencies measured through observation. There tended to be relatively large numbers of competencies loading on factors, especially factors extracted early in the analysis.

Although MANOVA provides descriptive information about the canonical variates, it is basically an inferential, statistics technique. Unlike the factor analysis, for the MANOVA groups of student teachers were identified, and the focus of interest was upon how these groups differed on the canonical variates. For these analyses, the number of possible canonical variates was considerably less than the number of possible factors from the factor analysis. Also, many more competencies loaded (with loadings greater than .30) on factors than on canonical variates. This fact produced differences between the characteristics of factors and canonical variates. However, such varied characteristics were expected since the groups of student teachers were quite similar in their performances on the competencies, and in order for competencies to load on canonical variates groups must differ on the competencies.

Summary

Since teaching is a behaviorally complex activity, multivariate analyses would appear to have potential for dealing with some of the complexity. Multivariate procedures have the advantage of being able to synthesize and "hold together" data rather than "taking apart" teaching performance piece by piece (competency by competency). From the analyses described in this chapter, some noteworthy points can be made.

The factor analysis, especially that of the COKER, indicated that identifiable constructs do underly teaching. The factors were quite distinctive and, of course, any objective criterion used for retaining factors will identify a finite number. The dominance of the first factor of the TPAI was somewhat disappointing.

The differences in the factor patterns for the TPAI and COKER data merit attention. Part of these differences was due to the different numbers of competencies measured by the two inventories; the TPAI measuring 45 versus 22 for the COKER, and the TPAI measuring supportive competencies which are not necessarily measured in classroom observations. However, there were 18 competencies measured by both inventories. Thus, the lack of similarity across the factor patterns is evidence of the instrument dependence of measurement in this area. Factor analysis may be a useful procedure for theory testing or *post hoc* description, but in either case it will be necessary to consider the type of observation used.

The MANOVAs were limited because of the small numbers of groups of student teachers and the similarity of the groups. Only the first canonical variate of the TPAI had a substantial number (9) of competencies loading on it. Several of these were competencies that could be demonstrated outside of the classroom, or at least did not require direct teaching. Again, only one competency, C-42 (Uses positive reinforcement patterns with students), loaded across both the COKER and TPAI, and for the TPAI, this competency had a modest loading. The classification for groups was not the same for the two analyses, which limits comparisons across two instruments; however, there was some overlap.

MANOVA would prove more useful if additional groups of teachers or student teachers were available and if these groups were quite diverse. In that case there would be more degrees of freedom so greater numbers of statistically significant, canonical variates could appear. Also, diversity among groups would tend to increase the numbers of competencies loading on canonical variates. MANOVA would be especially effective for synthesizing across-group differences, as the groups represent different programs, teaching areas and instructional situations.

Multivariate analyses will probably become increasingly useful as developmental work and research in observation continues. Such analyses aid in understanding the teaching-learning process. The analyses focusing on competencies individually as measured are certainly important, but multivariate analyses provide another perspective of the competencies. With the present lack of depth in the knowledge base of competent and effective teaching, it seems advisable to use any analyses that hold potential for explanatory power.

Chapter 6

OTHER RESULTS, CONSIDERATIONS, AND CONCLUSIONS

In preceding chapters we have reviewed and described the process of high and low inference observation of teaching and discussed the primary results from the analysis of such observation data. In this chapter, a number of related topics are discussed which includes projections of future use of observation data. The closing section provides present conclusions and recommendations based on three years of research/evaluation effort.

Competencies Measured by Both the COKER and the TPAI

There were 18, University of Toledo competencies measured by both observation instruments. In an earlier study,[47] scores for these 18 competencies on the TPAI and COKER were correlated, using the standardized T-score for scoring the COKER. The correlation coefficients in this study between the COKER and TPAI scores were modest; many were of zero magnitude, and only one (+.38) was statistically, significantly different from zero, using $\alpha = .05$ and a nondirectional test. With $N = 57$, a coefficient with an absolute value of .261 was required for significance. This result was somewhat disconcerting since one interpretation (and probably the most likely explanation) is that measurement in this area is highly instrument dependent.

The correlations between the TPAI score and the COKER scores, scored with both the proportionality score and the T-score[48] were computed for the Winter 1980 through Spring 1981 data. These correlation coefficients (columns r_1 and r_2) along with those of the earlier study (r_0) are presented in Table 10. Again, the correlations were of modest size and several negative coefficients appeared. The patterns were very similar for the two COKER scores.

---

[47] George E. Dickson and William Wiersma, RESEARCH AND EVALUATION IN TEACHER EDUCATION: A CONCERN FOR COMPETENT, EFFECTIVE TEACHERS. Toledo: The Center for Educational Research and Services, College of Education, The University of Toledo, May 1980. (ERIC, ED 194 540)

[48] Scores obtained using the proportionality model and the T-score model are highly correlated as would be expected.

Table 10

Correlation Coefficients[1] Between the COKER and TPAI Scores On 18 Competencies Measured by Both Inventories

$N_1 = 57, N_2 = 153$[2]

	Competency	r_0	r_1	r_2
C-16	Uses a variety of instructional strategies.	—.08	.27	.26
C-19	Establishes transitions and sequences in instruction which are varied.	.20	.16	.15
C-20	Modifies instructional activities to accommodate identified learner needs.	—.16	—.14	—.15
C-21	Demonstrates ability to work with individuals, small groups and large groups.	.02	.05	.04
C-22	Structures the use of time to facilitate student learning.	—.14	—.07	—.02
C-23	Uses a variety of resources and materials.	—.23	.03	.03
C-24	Provides learning experiences which enable students to transfer principles and generalizations to situations outside of school.	.08	.03	.15
C-31	Provides group communication (cooperation, interaction, learning from others) experiences for students.	.13	—.03	.00
C-32	Uses a variety of functional verbal and non-verbal communication skills with students.	.19	.11	.06
C-33	Gives clear directions and explanations.	.03	—.09	—.12
C-34	Motivates students to ask questions.	.12	.14	.10
C-36	Accepts varied student viewpoints and/or asks students to extend or elaborate answers or ideas.	.08	.05	.03
C-38	Provides feedback to learners on their performance.	.01	.34	.26
C-40	Maintains an envonment in which students are actively involved, working on-task.	.38	.13	.08
C-41	Implements an effective classroom management system for positive student behavior (discipline).	.08	.10	.08
C-42	Uses positive reinforcement patterns with students.	—.09	.19	.16
C-43	Assists students in discovering and correcting errors and inaccuracies.	.15	.20	.13
C-44	Develops student feedback, evaluation skills and student self evaluation.	.10	—.08	—.06

---

[1] r_0 is the correlation from the earlier study (1979-80), r_1 is the correlation between the TPAI and the proportionality score (1980-81), r_2 the correlation between the TPAI and T-score (1980-81).

[2] N_2 and N_2 are for the earlier study (1979-80) and the larger data analysis (1980-81), respectively.

The student teachers for whom the data were collected were not randomly selected from some larger population. So, the use of inferential statistics may be questioned. Yet an *a priori* argument could be made that these student teachers were representative of some larger populations, certainly the populations of student teachers going through the CBTE program at The University of Toledo. Also, the use of statistical significance gives some basis for judging the magnitude of a coefficient or a difference. This is similar to the rationale used by Medley in his review (1977),[10] in which he required a correlation coefficient to be statistically significant at $\alpha = .05$ for inclusion. Thus, the risk that a relationship due to chance would be retained is no greater than 5 percent.

With N = 153, a coefficient with an absolute value of .157 is required for significance. For the proportionality score six, or one-third of the correlations, were significantly different from zero, and for the T-score, three reached statistical significance. In all cases the significant, correlation coefficients were positive, certainly the desirable direction. These were non-directional tests using $\alpha = .05$. If a directional test is used, testing the alternate hypothesis that the correlation (in the population) is greater than zero, the hypothesis is supported for seven competencies using the proportionality score and for five competencies using the T-score (required value, .133). But even for those competencies having positive correlation coefficients, the sizes of these coefficients were modest, the largest being .34 on C-38, .30 on C-24 and .27 on C-16 for the proportionality score, and .26 on C-38 and C-16 for the T-score.

Another point to consider is the change in the correlations from the earlier analysis to the larger data analysis. The correlations between the proportionality score and the T-score for COKER data were generally around .90, and the correlation patterns with the TPAI are similar, as has already been noted. A difference of approximately .28 to .30 is required for statistical significance at $\alpha = .05$, nondirectional test.

There were few, significant changes in correlations across the two analyses. C-16 had a significant change with COKER scores on both scoring systems. C-38 and C-42 had significant differences for the proportionality score only. There were many coefficients for which the changes were very small, only a few one-hundredths of a point. In some instances the change went from positive to negative, in others from negative to positive.

---

[10] Donald M. Medley, Op Cit (see footnote on page 2).

Considering the three competencies for which the differences in coefficients were significant, two changed from a non-significant negative value, to a significant positive value (C-16 and C-42). The other competency, C-38, changed from .01 to a significant, positive value. So at least, the relationships for these competencies were going in the desired direction.

What can be concluded from these results? Overall, the results from the two analyses are quite consistent. It is unlikely that the general lack of relationships between the scores is due to chance or some unique characteristics of these analyses. The statistics are quite stable, especially those of the larger analysis with N = 153, and they are quite stable across the two analyses. It apparently can be concluded that the lack of relationship is due to the instruments not measuring the skills in the same manner.

This conclusion may not be all that surprising when considering the conceptual difference between low-inference and high-inference observation. High-inference scores are based on a qualitative scale; low-inference scores have only a quantitative occurrence or lack of occurrence of behaviors. Essentially, the results show that the scores are independent. Knowing how an individual scored for a competency on one observation system gives no indication of what the score would have been on the other system. At least there were no significant, negative correlations, which would have been a most undesirable result.

Although the correlations of scores across the two observation instruments on the same competencies generally revealed independence between high and low-inference observation data, a review of the positioning of group means as provided in the profiles indicated considerable consistency. In only four of the 18 competencies (C-21, C-24, C-36, C-42) were the profile directions clearly opposite when going from one instrument to the other. On the other 14 competencies, results across the instruments were quite similar. In other words, if the TPAI showed a competency to be rated high, the same result tended to occur on the COKER. Thus, it appears that for discriminating between individuals within groups or positioning student teachers in groups, the results from the two instruments tend to be independent. But if the overall mean for any group of student teachers on a given competency tends to be up (high) or down (low) using one of the instruments, it will likely be the same way for the other.

Probably the most likely explanation of these results is that measurement in this area is highly instrument dependent when using both low-inference and high-inference inventories. Although this is not a desirable characteristic, it is one with which measurement of teacher competencies will have to live. It has definite implications for those who would use observation data for certification purposes, or for any substantive decision making for that matter. Instruments or inventories should be carefully selected, based on what the potential user expects from the data. If the qualitative characteristic, which is subjective, is desired, a score including this characteristic will be independent of a score based on occurrence of behaviors, the latter being considered the more objective score. If a user's concept of teacher performance is "quantitative-based," the low-inference instrument score would be the most appropriate. In any event, this instrument dependence of the score should not be masked, because without further information the profession generally will assume that the same competency will be measured consistently across instruments.

The Use of Observation Data In Program Evaluation

The development work and the logistics of data collection connected with observation are not ends unto themselves, although at times they may seem that way. The payoff with observation is the use made of the results. However, it should not be inferred that the development work is unimportant, because it is extremely important and essential. Not only that, but with the present "state of the art," continued research and development in this area are necessary and will continue to be necessary for the foreseeable future.

There are a number of uses for observational data and some are discussed later in the chapter. The initial, continuing and primary purpose of the observation of University of Toledo student teachers is for program evaluation and individual assessment and diagnosis. Program evaluation is considered in this section followed by a discussion of individual evaluation concerns.

The observation/performance approach to program evaluation supplements the assessment of subject area and pedagogical knowledge. It is a different approach than evaluating programs by the quantification of college courses taken and credits accrued. One of the spinoff benefits of evaluation through observation is that it requires the identification of the competencies intended from the program. Who could argue against preparing competent teachers? But many teacher education programs operate at a level that does not specify competencies; at best they are implicit. There is value in the process of explicitly identifying competencies. It causes the faculty to reflect on what it is their students are supposed to be able to do upon completion of the program. In some respects, the identification of competencies is an alternate approach to program definition, alternate to the usual approach of specifying courses and credits. It is not likely that competencies will replace courses and credits, but they supplement the usual approaches to program definition.

As far as The University of Toledo is concerned, it has been demonstrated that competencies can be identified, and subsequently, student teacher performance on the competencies can be measured through observation. The student teachers demonstrate varying degrees of proficiency on the competencies intended from the CBTE program. Present observation results reveal a few weak areas; specifically those related to evaluation and student feedback, and to a lesser extent, some of the skills relating to the individualization of instruction. Evaluation procedures probably should receive stronger emphasis in the program. The University of Toledo program does emphasize individualization, especially at the elementary level with its Individually Guided Education (IGE) components. It is possible that some observers are highly sensitive to individualization competencies, expect more from the student teachers and correspondingly, tend to score them lower on these competencies. Other observers may be less concerned about individualization, especially at the secondary level, and gloss over such competencies. It would be informative if data were available on other groups of student teachers, observed by the same individuals.

Observation data are useful for program evaluation internal to a university or college. Evaluation across programs would also be informative, especially as relative strong and weak points could be identified. However, it is difficult to implement across-program evaluation, not only because of costs, but because of the necessary inter-university cooperation. Yet the costs are certainly not prohibitive considering the value of evaluation. If data were used for formative evaluation, which they should be, universities and colleges should not be threatened, and the necessary cooperation should be forthcoming. This type of effort could receive motivation and possibly initial development through the state agency for teacher education programs and certification.

Individual Diagnosis and Assessment

The profiles discussed in Chapter 4 consisted of group profiles; however, profiles can be generated for the individual student teacher or teachers. The model for individual diagnosis and assessment is one of comparing an individual's scores on one or more competencies with those of some normative group. The selection of the normative group merits consideration, and for student teachers the most appropriate group would probably be a previous group of student teachers at the close of the student teaching experience. A group of experienced, successful teachers would also make a relevant comparison group, its performance sort of an empirically based standard. However, such data may be difficult to obtain and there would need to be some type of agreement on the criteria for successful teachers. However, as the data base expands, additional, usable normative groups should become available.

The individual profiles can be used in consultation with the student teacher in diagnosing strengths and weaknesses of the teaching performance. Figures 6 and 7 contain individual profiles on the COKER and TPAI with a comparison group as generated through a computer terminal with graphics capability. Computer programs exist for obtaining such profiles for individual competency scores on both the COKER and the TPAI at The University of Toledo.

The present technology is such that student teachers could receive almost immediate feedback on the student teaching performance. However, the sporatic availability of computer terminals with graphic capabilities on The University of Toledo campus has not yet made immediate feedback feasible. Conceivably, the score from an observation could be processed through the terminal immediately after the observation is completed, and profiles such as those in Figures 6 and 7 could be obtained on the terminal screen. The supervisor and the

Profile of a Student's Competency Scores Compared to the Norm Group - COKER

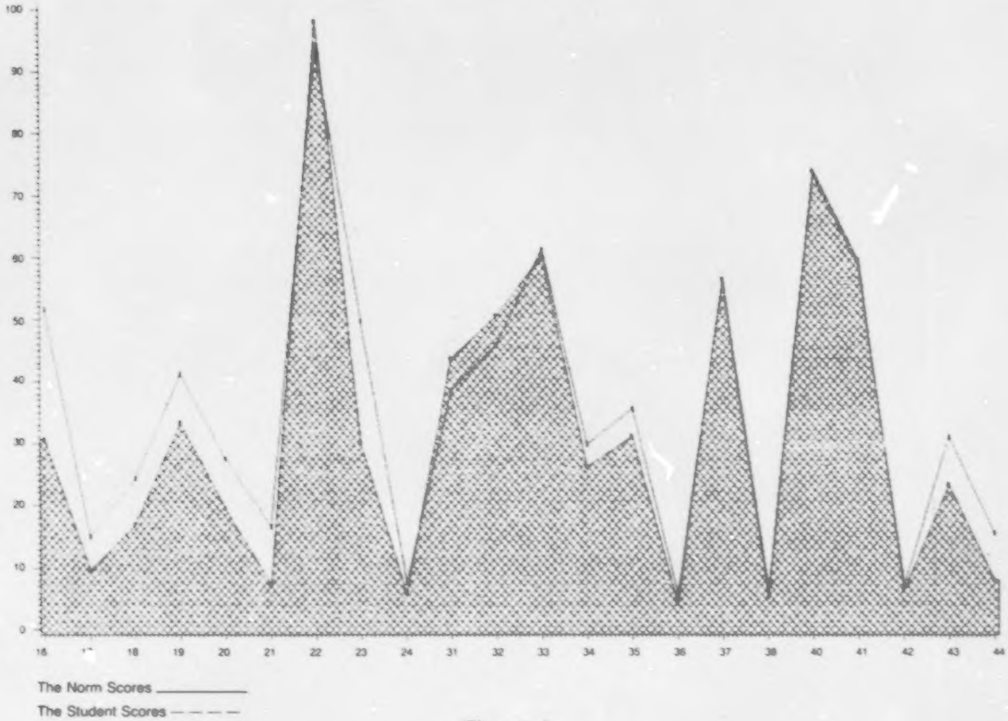


Figure 6

Profile of a Student's Competency Scores Compared to the Norm Group - TPAI

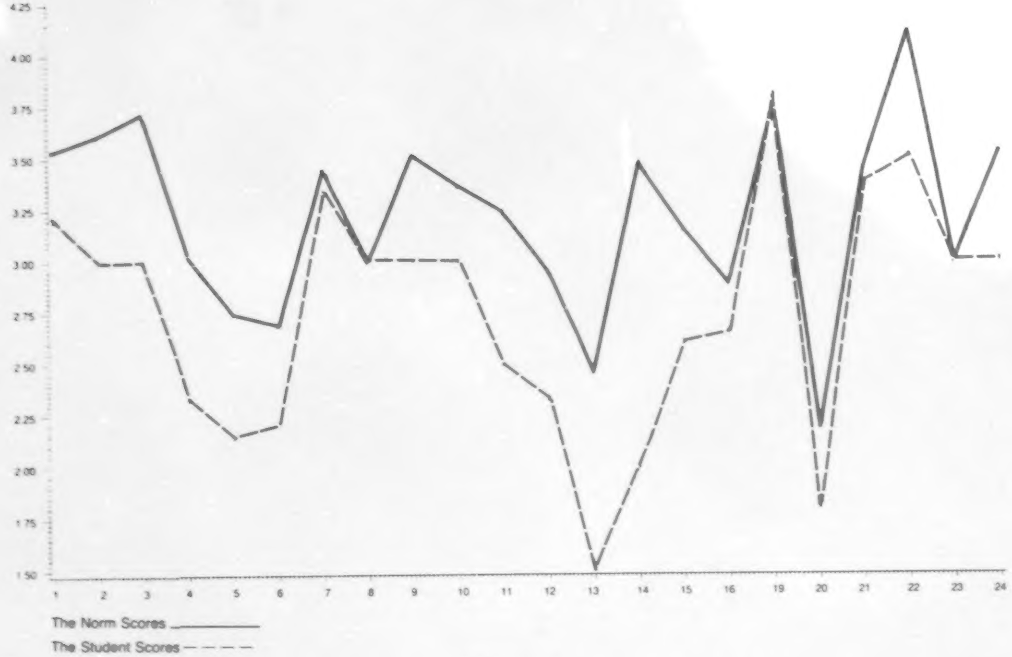


Figure 7

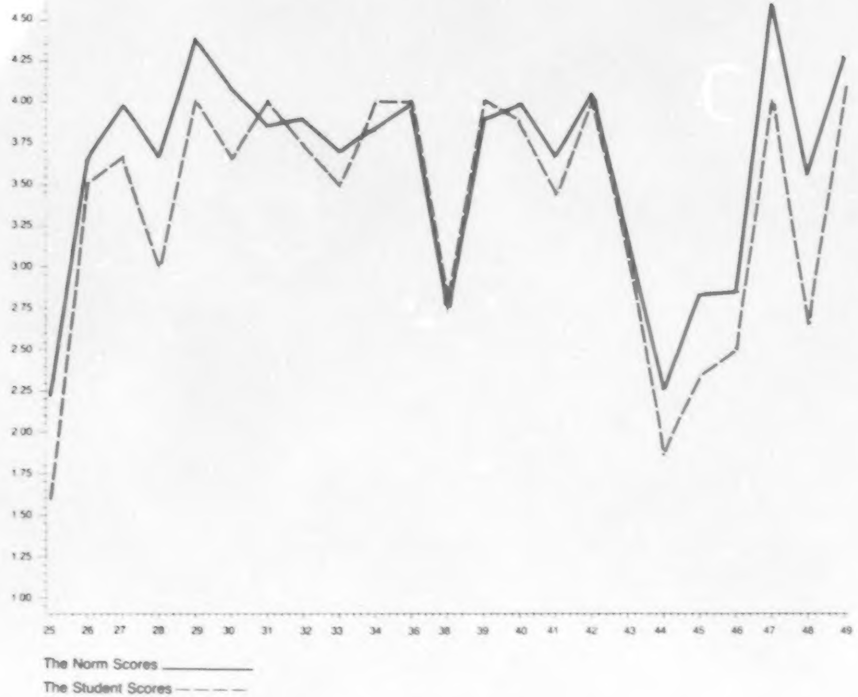


Figure 7 (cont'd.)

student teacher could then discuss the results soon after the observation occurred.

Presently, computer profiles can be generated for individual teachers across a single competency or for a single teacher across several competencies. Such profiles are illustrated in Figures 8 and 9. These figures are taken from the *COKER User's Manual*[50] and the scores used are standardized T-scores. The profiles are based on data from experienced teachers and the relative positioning is based internally on the group of 43 teachers. The purpose of the observation for the experienced teachers was for the diagnosis of individual strengths and weaknesses. The same use of individual profiles could be applied with student teachers.

Although observation data are useful for program evaluation, their use in the long term for individual diagnosis and assessment may prove the more valuable of the two uses. Such technology and capability are available, it is basically a matter of making the commitment to implement the process. This approach to feedback and evaluation of the student teaching performance provides for increased objectivity of the evaluation, and certainly could supplement, even supplant present evaluation procedures. Colleges of Education have a gatekeeping function to perform. Final approval of graduates for the teaching profession should be based more extensively, than it is presently, on classroom teaching behavior and competence.

Other Uses of Observation Data

The primary use of observation data as discussed in this report has been for program evaluation/research purposes. As such, considerable background, developmental and exploratory work needed to be done. For example, the proportionality model for scoring COKER data was developed. Factor analyses were done more as a study of the underlying constructs of teaching, than as a basis for making evaluation decisions. The entire effort was one of basically trying new procedures and instruments.

The state of the art of classroom observation is far enough along so that there are a number of uses, other than research purposes, that can

---

[50] Joan G. and Homer Coker, Op. cit., pp. 48-49, 51.

Figure 8
A Profile of Scores of 43 Teachers Across the Competency, "Maintaining Student Involvement, Time-On-Task"

Teacher No.	Mean
1401	64.83
1402	75.02
1403	41.05
1404	60.81
1405	42.76
1407	47.39
1408	39.92
1409	52.69
1410	36.43
1411	54.30
1412	50.47
1413	67.20
1414	69.34
1415	38.69
1416	64.97
1417	53.27
1418	47.45
1419	34.51
1420	50.50
1421	44.62
1422	47.44

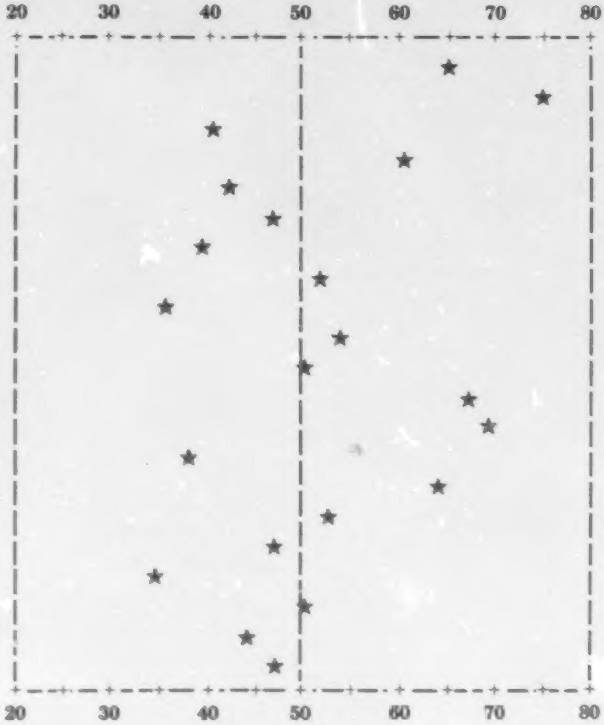


Figure 8 (Continued)

Teacher No.	Mean
1423	47.34
1424	50.91
1425	37.98
1426	54.12
1427	57.60
1428	35.81
1430	45.14
1431	42.50
1432	48.02
1433	52.65
1434	49.96
1435	66.90
1436	47.31
1437	42.17
1438	42.15
1439	40.56
1440	44.22
1441	55.17
1442	35.11
1443	55.47
1444	65.37
1445	49.88

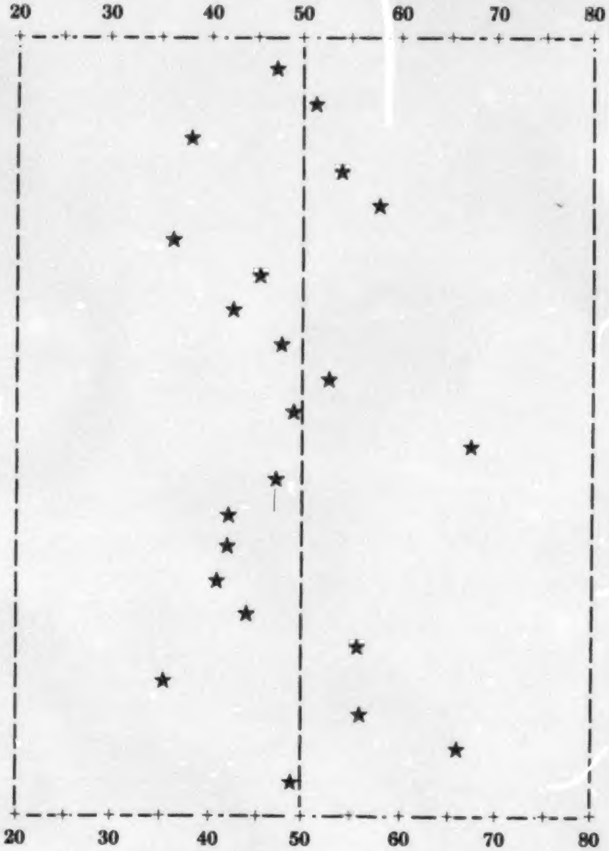
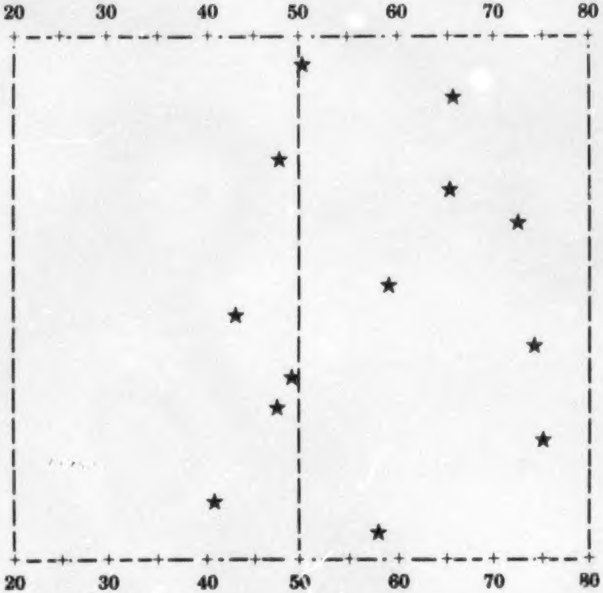


Figure 9

Profile for Teacher Number 1402 Across
13 Competencies or Behaviors

Competency	Score
1. Teachers work with large groups	50.10
2. Small groups work without adult supervision	65.93
3. Seatwork by S with T focus & structure & allow S choice of either How What When	48.36
4. Structures academic time	66.66
5. Students initiate verbal interactions	73.23
6. Student correct substantive responses to teachers questions	59.76
7. Low cognitive level questions	43.61
8. Teachers amplify and discuss student responses	74.41
9. Teacher hostility and rebuking behavior	49.91
10. Disruptive student behavior	47.35
11. Student involvement (time on task)	75.02
12. Non-substantive interaction between teacher and students	41.11
13. Structured student behavior	58.68



be applied in education settings. The emphasis in this report, so far, has been upon teacher education program research. Selected uses of observation data with inservice teachers are described below.

Assessment of Classroom Interaction

Observation can provide objective information about what is happening in the classrooms of a school or school system. There may be situations, for example those in which there are acute discipline problems or teacher morale is low, in which an objective record of classroom interaction would be helpful in developing a solution to the problem. In such situations there would be no intent of teacher assessment or evaluation. The use of the observation data would be to provide objective information so informed decisions could be made toward correcting the problem.

Inservice and Staff Development Programs

Inservice and staff development programs, however conducted, have an inherent purpose of improving the skills and knowledge of the teachers. Observation data on a group of teachers can provide information about the topics that might be most helpful in improving teacher effectiveness. Observation could be used as a method of needs assessment by observing teachers prior to implementing a staff development program. If inservice or staff development is individualized for teachers, individual teacher profiles can be used for diagnosis and correction of weaknesses.

Although observation, of necessity, focuses on teacher behavior and classroom interaction, spinoff benefits can materialize for administrators, supervisors and other educational professionals. Principals become increasingly sensitive to classroom needs and problems as they participate with the teachers in structuring a program of observation. Insight can be obtained which will enable the principal to be a more effective, educational leader. Such insight comes not only from the analysis of the observation data, but also from the process by which observation takes place. The process provides an opportunity for cooperative activity between administrators and teachers; an activity which deals directly with the educational program of the school. At the very least, observation will make the entire staff more knowledgeable about what is going on in the school.

Certification of Teachers

The Department of Education at the state level has the legal responsibility for the certification of teachers. Traditionally, initial certification upon graduation from college, has been based on the completion of approved programs, courses taken, or credits accrued. However, some states are moving toward various types of examinations for certification, and there have been studies of basing certification, at least in part, upon the teacher's performance in the classroom setting. A performance-based approach to certification appeals to legislators and to the general public.

Since it is undesirable to have evaluation and accountability requirements legislated without adequate input from the profession, observation as applied to certification merits attention. Teacher education institutions and state departments of education are most directly concerned with certification requirements. However, certification is also the concern of the schools, since in the final analysis, those teachers certified are employed by the schools. Any developmental effort within a state addressing the use of observation data for certification would require extensive cooperation from various units and sectors of the profession. This in itself would have a unifying effect on the profession in that state, and would tend to lessen the professional fragmentation and isolation that often exists in education.

Conclusions and Recommendations

The three years of collection and analysis of observation data have been demanding but rewarding in benefits and information. A number of conclusions can be drawn from the results and the process of evaluation through observation. These are summarized below.

It is possible to identify the competencies intended to be developed by a teacher education program for the training of prospective teachers. The identification process itself has value by providing increased definition and clarity of desired program outcomes. It requires teacher education professors to be much more specific about teacher education program components, especially the expected outcomes of teaching relative to teacher performance.

What is beneficial for teacher education faculty and students is often helpful for inservice teachers. The process of identifying competencies or groups of desired behaviors is potentially valuable in a staff development context for inservice teachers. Such effort causes professionals to deal with the questions of "What is expected?", and it clarifies and structures the expected outcomes of the teaching process. Presently this type of inservice education is rarely implemented at the school level.

The University of Toledo identified 49 competencies for its CBTE program. These appear to be adequately comprehensive, and they provide a workable number of competencies. Many of the competencies are quite broadly stated, a necessity in order to avoid excessive numbers of competencies. Yet, it is important to emphasize that the competencies are operationally defined by the cells of the COKER and the indicators and descriptors of the TPAI. These provide specificity to the competencies; just as achievement in mathematics may be quite general, it is operationally defined by the knowledge and skills necessary to solve the problems of a specific, mathematics test.

The use of the COKER and the TPAI provide viable approaches to measuring demonstrated or underlying competencies intended from a teacher education program. Although measurement through observation is demanding, it can be conducted and usable results obtained. These instruments appear to provide adequate measurement of the competencies, and confidence in the instruments has been developed and then they have been applied for use in the Toledo CBTE program. Further, we conclude that these instruments could be used, successfully, to measure competencies in other teacher education programs.

Profile analysis provides a variety of useful information, not only about relative performance of groups of student teachers, but also information about the nature of competencies. The means, medians and standard deviations are descriptive of the distributions of scores for competencies. The positioning of the COKER scores on the scale of measurement for the proportionality score is reflective of the types of behaviors indicative of the competencies. Comparisons between groups provide some differences reflective of program emphasis. With the TPAI score, the absolute value of the score can be compared to a "standard," which although subjectively established usually is based on the results of some normative groups. For any of the scores, comparisons can be made with scores of available normative groups. Although the COKER and TPAI results discussed in this paper are group results, individual profiles can also be derived and compared with group means for individual diagnosis and evaluation of teaching performance.

University of Toledo student teachers continue to score above the midpoint (3.0) on the TPAI on most of the competencies measured by this instrument. This indicates that the competencies established for them in the CBTE program are generally being demonstrated in student teaching. This demonstration is somewhat more extensive for special education and elementary teachers than secondary school teachers. Competencies involving the use of evaluation techniques and individualization activities with students apparently lack sufficient student teaching demonstration and suggest a program deficiency in The University of Toledo CBTE program. University of Toledo student teachers reveal particular teaching skills in certain competencies stressed in the CBTE program.

The COKER data definitely indicate that The University of Toledo student teachers demonstrate the competencies intended from the CBTE program. The exceptions are again competencies related to evaluation and the individualization of instruction which corresponds to similar findings with the TPAI results. However, the COKER data also reveal that students had low scores on C-36, "Accepts varied student viewpoints and/or asks students to extend or elaborate answers or ideas." Student teachers may score low on this competency because they are compulsive about teaching the planned lesson and are not sufficiently receptive to student viewpoints and comments. Teacher education programs should emphasize and develop teacher skills which make student teachers more receptive to student comment and viewpoints. The COKER data further indicated that scores were low on C-42, "Uses positive reinforcement patterns with students." This low score most likely reflects the fact that the observation period on the COKER is relatively short and does not allow a lot of opportunity for the observation of reinforcement behavior.

The factor analyses computed on both that TPAI and COKER data support the general contention that there are some underlying traits or constructs of teaching. The results for the COKER are more distinctive and there is less correlation between the COKER factors than between the factors of the TPAI. Even with the predominant first factor of the TPAI, it can be argued that the competencies did not collapse into a single, omnibus factor using either data base.

102

The results of the multivariate analysis procedures (MANOVAs) revealed a great deal of similarity across the groups of student teachers. The maximum number of statistically significant canonical variates appeared, but this result is undoubtedly a function of the high statistical power due to the large number of student teachers involved. Further, these variates had relatively few loadings with the competencies which is evidence of similarity of the groups. These characteristics were in some respects expected; however, since the groups of student teachers were quite similar in their performances on the competencies and in order for competencies to load on canonical variates, groups must differ on the competencies. The idea of generic competencies with respect to a specific teacher education program is reinforced by the MANOVA results.

The lack of correlation between scores on competencies measured by both the TPAI and the COKER indicates that measurement of teacher competencies through observation is highly instrument dependent, especially for individual scores. On the other hand, the consistency of group scores on the same competencies was encouraging. For all but four of the eighteen competencies measured by both instruments, the group scores across the instruments were consistent; that is, if the groups scored high on the compentency as measured by the TPAI, they also tended to score high on the same competency as measured by the COKER (and vice versa). Nevertheless, it appears that discrimination among scores within a group on the same competency is pretty much independent across the two instruments. This may be a characteristic of observation using high inference and low inference approaches, but it has important implications for interpretation and the use of results from such instruments. With this characteristic it is extremely important that potential users clearly understand the correspondence between the nature of the instruments and their objectives for using them.

The research reported in this monograph has emphasized group results, although the potential uses of results for individual diagnosis and evaluation have been mentioned. The model for individual diagnosis and evaluation is now in the development stage, and it appears to be clearly feasible for use with individual students in teacher education programs. The technology for this type of evaluation exists (especially in the use of graphic presentations with computers); it is basically a matter of implementing the process.

Both approaches to the observation of teacher behavior, low-inference and high-inference, have proved viable in the teacher education research attempted at The University of Toledo. It is important to note the characteristics of the two approaches when using observation for a given purpose. In no case is either approach "all things to all people." Considering the present state of the art and the extensive developmental effort in evaluation which has been conducted, as reported by this study, it would appear unwise to abandon either approach at this time. Nevertheless, the results have been definitive enough so that decisions can be made about the advantages of one or the other approach, given a defined purpose for conducting observation.

Where do we go from here with the measurement of teacher competence? Actually, the results discussed in this monograph are part of a continuing effort. Such effort should be viewed as developmental, at least in part, not an effort limited to the accumulation of empirical results, however important such results are for evaluation purposes.

Analyses that provide information about the nature of teaching would be useful to the profession, especially since this area is often lacking adequate theory. Additional factor analyses would provide information about constructs of teaching. Factor analyses for different groups of teachers, for example, elementary and secondary, may reveal differences in the underlying constructs. Whether constructs are similar or different, the facts are important and have definite implications for teacher education and the profession. Studying varying groups, such as student teachers compared with experienced teachers, should also be informative.

With differing groups of student teachers or teachers, MANOVAs would be likely to produce more distinctive differences on canonical variates. These results too, would be informative about the constructs of teaching if they differ across different groups of teachers. It may be that theory development about teaching would be more productive if launched from the empirical base provided by observation data.

The use of observation with inservice teachers holds potential for formative evaluation. Indeed, such observation is being used as part of the Educational Testing Service, Formative Evaluation Program.[51]

---

[51] Albert Benderson (ed.), "Teacher Competence," Focus, Vol. 10. Princeton, N.J.: Educational Testing Service, 1982. pp. 14-15.

Observation should not be used for "hiring and firing" of inservice teachers. It should be used with the attitude that "any teacher's skills can be improved." The ETS program is undoubtedly good, but it is not likely that any ETS or other similar program can service the entire profession. Professionals responsible for staff development in school systems should exert the leadership at the local level, with appropriate support from universities, state departments of education, and other agencies.

Teacher education today is faced with numerous important issues —competency testing, a negative, national image, low economic incentives and declining student enrollments to mention a few. Many issues present serious problems not only to teacher education, but to the entire profession. Education has always been hampered by a fragmented approach throughout the profession when developing solutions to problems. It should not be inferred that observation of teacher behavior and competence is a "cure-all" for the problems of education. But the use of observation in evaluation does provide the potential for a unified approach to problem solution. Teacher education institutions can remain autonomous as gatekeepers of the teaching profession yet work cooperatively on teacher evaluation. Since State Departments of Education have the legal responsibility for teacher certification and control many related funds, these state level agencies have an excellent opportunity for exerting leadership. Indeed, many states have already taken organized approaches to solving the problems of education. In some cases, the impetus came from within the profession; in others it was generated from outside, often through legislation. The former is by far the desired situation.

Finally, it is important that a teacher education faculty or an inservice staff realize that the type of effort described in this monograph is a longterm, cooperative effort. At a college of education it involves all faculty. Certainly faculty members have different functions, but all contributions are important. Correspondingly, if a state-level effort is launched, it should be cooperative among the teacher education institutions and other educational agencies as appropriate. For example, follow-up observation of beginning teachers would require the cooperation of the school systems in which the teachers are employed. But whatever the case, observation should be viewed as a multi-year effort.

The past three years at The University of Toledo in attempting to measure teacher competencies through observation of student teaching at all levels have proven to be professionally exciting and rewarding. The course we are charting has considerable potential for the improvement of teacher education and teacher effectiveness.

APPENDICES

APPENDIX A

University of Toledo Competencies Assessed Using the TPAI and COKER Instruments and Competency Symbol Indicating Such Assessment

	Competency Statement	TPAI Symbol	COKER Symbol
Topic:	Planning, Teaching Materials/Equipment and Evaluation.		
1.	Plans units of instruction.	C-1	None
2.	Plans instruction at a variety of cognitive levels.	C-2	None
3.	Can state pupil outcomes and/or student course objectives in behavioral terms (behavioral objectives.)	C-3	None
4.	Has realistic expectations for the learning process and student readiness for learning.	C-4	None
5.	Gathers, organizes, and evaluates pertinent information about students for effective instruction.	C-5	None
6.	Identifies and evaluates learning problems of students in content area being taught.	C-6	None
7.	Keeps informed of current professional/subject area literature and curricular learning materials/resources available.	C-7	None
8.	Knows how to select (or construct), organize and use appropriate instructional materials and equipment to facilitate learning activities.	C-8	None
9.	Uses criteria and effective procedures for determining pupil achievement of learning objectives.	C-9	None
10.	Selects/develops appropriate assessment techniques and instruments for instructional activities.	C-10	None
11.	Collects, quantifies, and interprets data from appropriate assessment instruments.	C-11	None

	Competency Statement	TPAI Symbol	COKER Symbol
12.	Maintains evaluation records.	C-12	None
13.	Engages in professional development by obtaining and analyzing evaluative information concerning the effectiveness of instruction.	C-13	None
14.	Uses information about the effectiveness of instruction to revise it, with possible curriculum modifications.	C-14	None
15.	Relates to accountability issues concerning responsibilities to students, parents, and the instructional process.	C-15	None
Topic:	Instructional Strategies, Techniques, and/or Methods.		
16.	Uses a variety of instructional strategies.	C-16	C-16
17.	Uses convergent and divergent inquiry strategies.	None	C-17
18.	Develops and demonstrates problem solving skills.	None	C-18
19.	Establishes transitions and sequences in instruction which are varied.	C-19	C-19
20.	Modifies instructional activities to accommodate identified learner needs.	C-20	C-20
21.	Demonstrates ability to work with individuals, small groups and large groups.	C-21	C-21
22.	Structures the use of time to facilitate student learning.	C-22	C-22
23.	Uses a variety of resources and materials.	C-23	C-23
24.	Provides learning experiences which enable students to transfer principles and generalizations to situations outside of school.	C-24	C-24
25.	Provides assignments/learning opportunities interesting and appropriate to different ability levels of pupils.	C-25	None
26.	Demonstrates knowledge in the subject areas.	C-26	None

	Competency Statement	TPAI Symbol	COKER Symbol
27.	Demonstrates self-direction and conveys the impression of knowing what to do and how to do it.	C-27	None
28.	Works effectively as a member of an instructional team.	C-28	None
29.	Uses acceptable written and oral expression with learners.	C-29	None
30.	Adjusts components of the physical/learning environment over which the teacher has control to facilitate learning.	C-30	None
Topic:	Communication With Learners.		
31.	Provides group communication (cooperation, interaction, learning from others).	C-31	C-31
32.	Uses a variety of functional verbal and non-verbal communication skills with students.	C-32	C-32
33.	Gives clear directions and explanations.	C-33	C-33
34.	Motivates students to ask questions.	C-34	C-34
35.	Uses questions that lead students to analyze, synthesize and think critically.	None	C-35
36.	Accepts varied student viewpoints and/or asks students to extend or elaborate answers or ideas.	C-36	C-36
37.	Demonstrates proper listening skills.	None	C-37
38.	Provides feedback to learners on their cognitive performance.	C-38	C-38
39.	Expresses a positive personal attitude toward the teaching profession.	C-39	None
Topic:	Learner Reinforcement - Involvement.		
40.	Maintains an environment in which students are actively involved, working on-task.	C-40	C-40
41.	Implements an effective classroom management system for positive student behavior (discipline.)	C-41	C-41

	Competency Statement	TPAI Symbol	COKER Symbol
42.	Uses positive reinforcement patterns with students.	C-42	C-42
43.	Assists students in discovering and correcting errors and inaccuracies.	C-43	C-43
44.	Develops student feedback, evaluation skills and student self-evaluation.	C-44	C-44
Topic:	Professional Standards.		
45.	Accepts responsibility, is dependable.	C-45	None
46.	Evidences cooperation with others (teachers, administrators, support staff, parents, etc.) in planning and teaching.	C-46	None
47.	Acts as an appropriate model in terms of ethics, attitudes and values.	C-47	None
48.	Attends teacher and other professional meetings.	C-48	None
49.	Understands and follows school law, policies, and procedures and their effects on teachers, and teaching, including professional conduct standards.	C-49	None

APPENDIX B

The Generic Competencies of Effective Teaching as Related to the Ohio Standards for Colleges and Universities Preparing Teachers

	Competency	Mean Rating On a 3.0 Scale
1.	Gives clear directions and explanations.	2.81
2.	Evidences fairness, tact, compassion and good judgment in dealing with pupils.	2.73
3.	Demonstrates knowledge in the subject areas.	2.68
4.	Gives student individual help or attention.	2.67
5.	Provides opportunities for all ability levels of pupils to respond and participate.	2.66
[illegible]	Demonstrates enthusiasm for teaching and learning and for the subject being taught at the time.	2.66
7.	Maintains an educational environment conducive to developing positive attitudes toward learning.	2.65
8.	Uses effectively a variety of verbal and non-verbal classroom communication techniques.	2.64
9.	Maintains a social classroom atmosphere which reflects enthusiasm, warmth, support, and respect.	2.64
10.	Maintains self-control in classroom situations with pupils.	2.63
11.	Controls disruptive or deviant pupil behavior objectively.	2.61
12.	Selects goals and objectives appropriate to pupil needs.	2.60
13.	Prepares lessons that are well organized and cohesive.	2.59
14.	Promotes self-awareness and positive self-image in pupils.	2.58
15.	Modifies instruction appropriate to identified learner needs.	2.58
16.	Accepts responsibility.	2.57

	Competency	Mean Rating On a 3.0 Scale
17.	Encourages students to take responsibility for their own work.	2.55
18.	Uses acceptable written and oral expression with learners.	2.55
19.	Demonstrates ability to work with individuals, small groups, and large groups.	2.54
20.	Identifies and evaluates learning problems of students in content area being taught.	2.54
21.	Uses positive reinforcement patterns with students.	2.53
22.	Provides accurate and prompt feedback to learners about their performance.	2.53
23.	Employs a variety of appropriate instructional strategies and techniques to achieve objectives.	2.52
24.	Makes possible an environment where students are actively involved, working on-task.	2.52
25.	Has realistic expectations for student learning.	2.52
26.	Knows how to select, organize and evaluate appropriate instructional materials.	2.51
27.	Initiates instruction by providing appropriate background information and explanation of task-related activities.	2.48
28.	Maintains a challenging level of instruction.	2.47
29.	Uses skillful questions that lead pupils to analyze, synthesize and think critically.	2.45
30.	Modifies the curriculum to serve changing needs of students.	2.44
31.	Uses valid criteria and procedures for determining pupil achievement of learning objectives.	2.44
32.	Structures the use of time to facilitate learning and student attainment of educational objectives.	2.43
33.	Provides feedback to student on his/her behavior.	2.43

	Competency	Mean Rating On a 3.0 Scale
34.	Expresses humor when appropriate.	2.42
35.	Selects, identifies and implements instructional procedures which are consistent with instructional objectives.	2.41
36.	Provides assignments which are interesting and appropriate to different ability levels.	2.41
37.	Motivates students to ask questions.	2.40

114

APPENDIX C

Classroom Observations Keyed for Effectiveness Research

115

CLASSROOM OBSERVATIONS KEYED for EFFECTIVENESS RESEARCH

1981

SUBJECT	CODES	DISTRICT	SCHOOL	TEACHER	CLASS SIZE	DATE MO	DATE DAY	DATE YEAR	RECORDER	GRADE	VISIT
⓪ ⓪	01 – Music & Art	⓪ ⓪ ⓪	⓪ ⓪ ⓪	⓪ ⓪ ⓪ ⓪	⓪ ⓪	⓪ ⓪	⓪ ⓪	⓪	⓪ ⓪ ⓪	[illegible] ⑦	⓪
① ①	02 – Lang. Arts	① ① ①	① ① ①	① ① ① ①	① ①	① ①	① ①	①	① ① ①	① ⑧	①
② ②	03 – Literature	② ② ②	② ② ②	② ② ② ②	② ②	②	② ②	②	② ② ②	② ⑨	②
③ ③	04 – Math	③ ③ ③	③ ③ ③	③ ③ ③ ③	③ ③	③	③ ③	③	③ ③ ③	③ ⑩	③
④ ④	05 – Reading	④ ④ ④	④ ④ ④	④ ④ ④ ④	④ ④	④	④	④	④ ④ ④	④ ⑪	④
⑤ ⑤	06 – Science	⑤ ⑤ ⑤	⑤ ⑤ ⑤	⑤ ⑤ ⑤ ⑤	⑤ ⑤	⑤	⑤	⑤	⑤ ⑤ ⑤	⑤ ⑫	⑤
⑥ ⑥	07 – Soc. Stu.	⑥ ⑥ ⑥	⑥ ⑥ ⑥	⑥ ⑥ ⑥ ⑥	⑥ ⑥	⑥	⑥	⑥	⑥ ⑥ ⑥	⑥	⑥
⑦ ⑦	08 – Planning	⑦ ⑦ ⑦	⑦ ⑦ ⑦	⑦ ⑦ ⑦ ⑦	⑦ ⑦	⑦	⑦	⑦	⑦ ⑦ ⑦		⑦
⑧ ⑧	09 – Transition	⑧ ⑧ ⑧	⑧ ⑧ ⑧	⑧ ⑧ ⑧ ⑧	⑧ ⑧	⑧	⑧	⑧ ⑧	⑧ ⑧ ⑧	[illegible] [illegible]	⑧
⑨ ⑨	10 – Other	⑨ ⑨ ⑨	⑨ ⑨ ⑨	⑨ ⑨ ⑨ ⑨	⑨ ⑨	⑨	⑨	⑨ ⑨	⑨ ⑨ ⑨		⑨

SECTION A

	STUDENT	1. Passive compliance	2. Active compliance	3. Answers quest. directed to him	4. Answers quest. voluntarily	5. Substantive question	6. Procedural question	7. Volunteers comment	8. Off-Task Questions or comments	9. Individual Off-Task (V or Non V)	10. Disruptive Off-Task (V or Non V)
PRESENTING	11 Orienting	○	○	○	○	○	○	○	○	○	○
	12 Motivates/Prob. struct. (set induction)	○	○	○	○	○	○	○	○	○	○
	13 Directs – learning related	○	○	○	○	○	○	○	○	○	○
	14 Directs – non learning related	○	○	○	○	○	○	○	○	○	○
	15 Explains, discusses, tells	○	○	○	○	○	○	○	○	○	○
	16 Digresses	○	○	○	○	○	○	○	○	○	○
	17 Praises	○	○	○	○	○	○	○	○	○	○
QUESTIONING	18 Open ended/No wrong answer	○	○	○	○	○	○	○	○	○	○
	19 Recall/rote/information	○	○	○	○	○	○	○	○	○	○
	20 Use or Application	○	○	○	○	○	○	○	○	○	○
	21 Ampli/Eval/Elaborate – Students' own idea	○	○	○	○	○	○	○	○	○	○
	22 Ampli/Eval/Elaborate – Other Students' idea	○	○	○	○	○	○	○	○	○	○
	23 Status	○	○	○	○	○	○	○	○	○	○
RESPONDING	24 Praises with explanation	○	○	○	○	○	○	○	○	○	○
	25 Praises without explanation	○	○	○	○	○	○	○	○	○	○
	26 Tells, gives info	○	○	○	○	○	○	○	○	○	○
	27 Wrong answer/No answer/Gives info	○	○	○	○	○	○	○	○	○	○
	28 Accepts – neutral/checks own perception	○	○	○	○	○	○	○	○	○	○
	29 Uses, extends	○	○	○	○	○	○	○	○	○	○
	30 Waits (cog)	○	○	○	○	○	○	○	○	○	○
	31 Focusing/Cueing	○	○	○	○	○	○	○	○	○	○
	32 Asks another student	○	○	○	○	○	○	○	○	○	○
	33 Repeats after feedback	○	○	○	○	○	○	○	○	○	○
	34 Listens/Observes/Pays attention	○	○	○	○	○	○	○	○	○	○
	35 Suggests/Questions/Directs/Ignores	○	○	○	○	○	○	○	○	○	○
	36 Interrupts/Rejects	○	○	○	○	○	○	○	○	○	○
	37 Criticizes/Commands	○	○	○	○	○	○	○	○	○	○
	38/39 Code Involvement								① ② ③ ④ ⑤	☒	① ② ③ ④ ⑤

KEY
1 – less ¼ of class
2 – ¼ of class
3 – ½ of class
4 – ¾ of class
5 – all of class

Joan G. & Homer Coker
P.O. Box 1017
Carrollton, Georgia 30117

DO NOT WRITE IN THIS AREA

03066

SECTION B

STUDENT

○	40 Enthusiastic	
○	41 Praises another	
○	42 Pats, hugs Pos. horseplay	
○	43 Laughs, smiles	
○	44 Shows pride	
○	45 Agrees V or NV	
○	46 Friendly V or NV	
○	47 Helps/Shares/Consider V or NV	
○	less ¼ of class	48 Code Involvement (40-47)
○	¼ of class	
○	½ of class	
○	¾ of class	
○	all of class	
○	49 Intense Involvement	
○	50 Leadership	
○	51 Follows Routine WO Rem.	
○	52 S. answers another S	
○	53 Works w social	
○	54 Collab. work/play	
○	55 Compet. work/play	
○	56 Confused	
○	57 Self directed, inappro.	
○	58 Wanders about	
○	59 Pouts, withdraws	
○	60 Shows fear, shame	
○	61 Makes face, frowns	
○	62 Tattles	
○	63 Teases	
○	64 Resists	
○	65 Picks at another/Neg. horseplay	
○	66 Dem/Com/Boss/Con	
○	67 Dis/Annoy/Interferes	
○	68 Takes/Damage/Stamp/Throws	
○	69 Crit/Disparages	
○	70 Att/Hit/Hurt/ - agg	
○	less ¼ of class	71 Code Involvement (61-70)
○	¼ of class	
○	½ of class	
○	¾ of class	
○	all of class	
○	less ¼ of class	72 Code Interest/Attention
○	¼ of class	
○	½ of class	
○	¾ of class	
○	all of class	

TEACHER

METHODOLOGY		
Motivation	○	73 Intrinsic Immed
	○	74 Intrinsic Future
	○	75 Extrinsic
Student Choice	○	76 S no choice
	○	77 S lim choice
	○	78 S free choice
Super vision	○	79 Aloof, detached
	○	80 Observes, Monitors
	○	81 Joins, partic
	○	82 Manages simul. Act
	○	83 Close superv
Focus	○	84 T Q Prob
	○	85 S Q Prob
Source	○	86 Text, pkg. res
	○	87 Multiple res
Diver/Evaluation	○	88 Same mat/eval
	○	89 Indiv. mat/eval
	○	90 S. partic. eval
Student Expres	○	91 Discourages
	○	92 Encourages
Student Perpi	○	93 Prevents
	○	94 Fosters
Mis inform	○	95 Accepts
	○	96 Corrects
Strategy	○	97 Inductive
	○	98 Deductive
	○	99 Transductive
	○	100 Expository
Cogn Level	○	101 Simple
	○	102 Complex
Use of Student time	○	103 T. assigns SA
	○	104 Indep. work

AFFECT	NON VERBAL	CONTROL	
○	106 Warm, cong		
○	107 Nod, smiles	○	108
○	109 Tou/Pat/Hug	○	110
○	111 Pause	○	112
○	113 Eye contact	○	114
○	115 Ignore	○	116
○	117 Gestures/Sig./Raps	○	118
○	119 Shakes head/shh	○	120
○	121 Takes something	○	122
○	123 Glares, Frowns	○	124
○	125 Holds, Pushes	○	126

VERBAL

○	127 Praise non sub	○	128
○	129 Agrees, Supports	○	130
○	131 Pos. Indiv. Attn	○	132
○	133 Reminds	○	134
○	135 Says stop	○	136
○	137 Firm/Sharp	○	138
○	139 Sco/Warns/Pun	○	140

GROUPING (Blacken ◉ for dominant mode)

	Non pres			No of Stu	Prescribed			
	WT		WOT		WT		WOT	
141	○ ◉	146	○ ◉	1	○ ◉	150	○ ◉	155
142	○ ◉	147	○ ◉	2-3	○ ◉	151	○ ◉	156
143	○ ◉	148	○ ◉	4	○ ◉	152	○ ◉	157
144	○ ◉	149	○ ◉	+	○ ◉	153	○ ◉	158
145	○ ◉			ALL	○ ◉	154	○ ◉	159

COMMENTS:

DO NOT WRITE IN THIS AREA